Journal of Apiculture
2026.41(2) : 89-115
DOI: 10.17519/apiculture.2026.06.41.2.89

Original research article

High-quality Genome Assemblies of Seven Korean Honeybee Breeding Lines

Jeong-Hyeon Lee[1,2], Ji Ran Jeong[3], Olga Frunze[1,2], Hyun Jo Koo[3], Dongwon Kim[4], Sangmi Han[4], Gyoungju Nah[3,*] and Hyung Wook Kwon[1,2,*]

[1]Convergence Research Center for Insect Vectors, Incheon National University, Incheon 22012, Republic of Korea
[2]Division of Life Sciences, Incheon National University, Incheon 22012, Republic of Korea
[3]Genome Analysis Center, National Instrumentation Center for Environmental Management, Seoul National University, Seoul 08826, Republic of Korea
[4]Department of Agricultural Biology, National Institute of Agriculture Science, Rural Development Administration, Wanju 55365, Republic of Korea

Abstract Honeybees (*Apis* spp.) are indispensable pollinators, yet their genetic integrity is increasingly threatened by habitat fragmentation and genetic introgression. High-quality chromosomal blueprints for specific, managed breeding strains are urgently required to facilitate precision molecular breeding and robust germplasm conservation. In this study, we generated chromosome-level *de novo* genome assemblies for seven honeybee breeding lines registered in the Korean National Seed Bank: five *Apis mellifera* (Italian, Carpatica, and Carniolan origins) and two *Apis cerana*. Using a hybrid sequencing strategy combining high-depth PacBio HiFi long reads and Oxford Nanopore Pore-C proximity-ligation data, an optimized NextDenovo and NextPolish pipeline successfully anchored contigs into 16 chromosome-level scaffolds. The finalized assemblies ranged from 207.68 to 223.22 Mb in size with exceptional BUSCO completeness (98.1–98.2%), effectively eliminating the pseudo-haplotypic inflation common in alternative assemblers. Comparative functional genomics revealed a clear species-level dichotomy: *A. cerana* genomes were significantly enriched in structural and regulatory categories, including cellular anatomical structure and biological regulation, whereas *A. mellifera* lines exhibited pronounced expansions in catalytic, transporter, and ATP-dependent activities. At the intra-species level, the identification of a toxicological outlier (Am-V) was resolved as a composite signature of host defensive peptides and co-assembled endosymbiont sequences, rather than a broad genomic expansion. Concurrently, complete mitogenome mapping resolved maternal phylogeographic structures, anchoring the native *A. cerana* lines within the mainland East Asian continental clade and partitioning *A. mellifera* lines within the European C-lineage. These high-fidelity reference templates establish a critical digital infrastructure for high-throughput marker development and climate-adaptive apicultural genomics.

Keywords *Apis mellifera*, *Apis cerana*, Chromosome-level assembly, PacBio HiFi, Pore-C, Korean honeybee, Seed bank, Precision breeding

INTRODUCTION

Honeybees are among the most economically and ecologically important insect pollinators worldwide, contributing an estimated $200 billion annually to global crop production through pollination services (Klein *et al*., 2007). The Western honeybee (*Apis mellifera*) and the Eastern honeybee (*Apis cerana*) are the two most widely

Received 29 May 2026; Revised 17 June 2026; Accepted 19 June 2026
Jeong-Hyeon Lee and Ji Ran Jeong are co-first authors and contributed equally.
*Co-corresponding authors. E-mail: gnah.nicem@snu.ac.kr, hwkwon@inu.ac.kr

managed species, with *A. mellifera* dominating commercial apiculture in Europe, the Americas, and parts of Asia, while *A. cerana* remains the primary managed species in many East and Southeast Asian countries (Crane, 1999). Both species face mounting pressures from habitat fragmentation, pesticide exposure, parasites (notably *Varroa* mites), pathogens (including viruses and *Nosema* spp.), and climate change (Meixner, 2010; Goulson *et al*., 2015). These stressors have driven significant colony losses and heightened concerns about the long-term sustainability of apiculture and pollination-dependent agriculture (Potts *et al*., 2010).

Genetic diversity is a cornerstone of species resilience and adaptive potential (Frankham *et al*., 2010). In honeybees, natural subspecies and locally adapted ecotypes harbor allelic variation that underpins traits such as disease resistance, foraging efficiency, overwintering success, and behavioral adaptations to regional climates (Ruttner, 2013; Wallberg *et al*., 2014). However, intensive breeding practices, global trade in queens and colonies, and inadvertent hybridization have led to genetic homogenization and the erosion of native gene pools in many regions (De la Rúa *et al*., 2009; Harpur *et al*., 2012). The conservation of genetically distinct breeding lines is therefore critical not only for preserving biodiversity but also for maintaining a reservoir of adaptive alleles that can be leveraged in breeding programs aimed at improving honeybee health and productivity (Meixner *et al*., 2010).

In South Korea, apiculture has a long history, with both introduced *A. mellifera* subspecies (primarily Italian and Carniolan bees) and native *A. cerana* populations playing important roles in agriculture and ecosystem services (Jung and Lee, 2018). Recognizing the need to safeguard these genetic resources, the Korean government has established a National Seed Bank program that includes the registration and maintenance of core honeybee strains under controlled, genetically isolated conditions (Frunze *et al*., 2022). These strains are maintained at specialized facilities such as the Wido Isolated Breeding Station in Buan-gun, Jeollabuk-do Province, which provides geographic isolation to minimize genetic contamination from feral or commercial colonies (Frunze *et al*., 2022). Despite the importance of these conserved strains, comprehensive genomic characterization has been lacking. While high-quality reference genomes have been established domestically for the native *A. cerana*—ranging from the early next-generation draft (AcSNUv2.0) to the recent near-complete chromosome-level layout (Acer K1.0)—there is currently no domestic reference assembly available for *A. mellifera* strains maintained in South Korea, enforcing a total reliance on generic foreign templates (Park *et al*., 2015; Lee *et al*., 2025). Chromosome-level genome assemblies of multiple distinct strains for both species are therefore essential for comparative genomics, accurate identification of line-specific structural rearrangements, and the systematic development of molecular markers tailored for precision breeding (Chapman *et al*., 2015; Wragg *et al*., 2016).

Recent advances in long-read sequencing technologies have revolutionized genome assembly. PacBio High Fidelity (HiFi) sequencing generates long reads (10–25 kb) with exceptionally high accuracy (>99.9%, Q20 or higher), enabling the resolution of repetitive regions and the production of highly contiguous assemblies with minimal errors (Wenger *et al*., 2019; Cheng *et al*., 2021). Complementary scaffolding approaches, such as Hi-C and related proximity-ligation methods, use chromatin conformation capture to infer long-range genomic contacts, facilitating the ordering and orientation of contigs into chromosome-scale scaffolds (Lieberman-Aiden *et al*., 2009). Oxford Nanopore Technologies (ONT) Pore-C is a recent innovation that extends traditional Hi-C by capturing multi-way chromatin contacts, providing richer information for scaffolding and enabling chromosome-level assembly even in complex or polyploid genomes (Goel *et al*., 2019; Garg *et al*., 2021).

In this study, we present chromosome-level genome assemblies for seven honeybee strains registered in the Korean National Seed Bank: five *A. mellifera* strains representing Italian, Carpatica/Carniolan, and Carniolan lineages, and two native *A. cerana* strains from Korea. We employed a hybrid sequencing strategy combining Pac Bio HiFi long reads for contig assembly and ONT Pore-C data for chromosome-level scaffolding. We assessed assembly quality using standard metrics (N50, BUSCO completeness) and Hi-C contact maps, and we performed phylogenetic analysis based on mitochondrial genomes to confirm the genetic identity and evolutionary relationships of these strains. Our results provide a high-quality genomic foundation for the conservation, management, and molecular breeding of Korean honeybee genetic resources.

MATERIALS AND METHODS

1. Sample collection and preservation

Seven honeybee strains were selected from the Korean National Seed Bank, comprising five *A. mellifera* strains and two *A. cerana* strains. The *A. mellifera* strains included three of Italian origin (Am-A, Am-C, Am-F), one Carpatica/Carniolan-type strain (Am-D), and one Carniolan strain (Am-V). The *A. cerana* strains (Ac-R, Ac-X) were both of Korean origin. For each strain, more than 50 individual worker bees were collected to ensure adequate representation of colony genetic diversity. Samples were collected from two primary locations in Jeollabuk-do Province, South Korea: the Wido Isolated Breeding Station in Buan-gun and the National Institute of Agricultural Sciences in Jeonju. The Wido station is situated on an island, providing geographic isolation that minimizes the risk of genetic contamination from external colonies. All samples used for genomic analyses were nurse bees aged approximately 3–7 days post-eclosion, collected directly from the brood nest area of each colony. Immediately upon collection, specimens designated for DNA extraction were individually flash-frozen using portable liquid nitrogen dry shippers and transported to the laboratory, where they were stored at −80°C until DNA extraction.

2. High Molecular Weight (HMW) DNA extraction

HMW DNA was extracted from whole body of individual nurse bees using a modified CTAB (cetyltrimethylammonium bromide) protocol optimized for insect tissues. Briefly, whole body were homogenized in liquid nitrogen, incubated in CTAB extraction buffer (2% CTAB, 1.4 M NaCl, 100 mM Tris-HCl pH 8.0, 20 mM EDTA, and 1% beta-mercaptoethanol) at 37°C for 1 hour, followed by chloroform:isoamyl alcohol (24 : 1) extraction and ethanol precipitation. DNA pellets were washed with 70% ethanol, air-dried, and resuspended in TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). Extract quality was assessed by 1% agarose gel electrophoresis and quantified with the Quantus Fluorometer (Promega, USA).

3. PacBio HiFi sequencing

The long read libraries were prepared using the Pac Bio SMRTbell Prep Kit 3.0 and SMARTbell barcoded adapter plate 3.0 according to the manufacturer's protocol (Pacific Biosciences). Briefly, 5 μg of HMW DNA per sample was sheared to a target size of 15–20 kb using a g-TUBE (Covaris), followed by DNA damage repair, end repair, and ligation of SMRTbell adapters. Size selection was performed using the PippinHT system (Sage Science) to enrich for fragments >10 kb. Library quality and concentration were assessed using the Femto Pulse systems (Agilent) and Quantus Fluorometer. To maximize sequencing efficiency and cost-effectiveness, libraries from 7–8 individual honeybee strains were pooled in equimolar ratios and loaded onto a single SMRT Cell. Long-read sequencing was performed on the PacBio Revio system using Sequencing Kit 3.0 and the 30-hour movie acquisition protocol. This pooling strategy was designed to achieve approximately 20× genome coverage per strain, assuming a genome size of approximately 230 Mbp for *A. mellifera* and 220 Mbp for *A. cerana*. Circular Consensus Sequencing (CCS) was performed using SMRT Link v13.3 software with default parameters to generate HiFi reads with a minimum predicted accuracy of Q20 (99% accuracy).

4. Oxford Nanopore Pore-C sequencing

Pore-C sequencing based on the Oxford Nanopore sequencing platform was employed to acquire chromosomal contact information for insect tissues. For each strain, three whole nurse bees were fixed with 1% formaldehyde for 10 minutes at room temperature to crosslink chromatin, followed by quenching with 125 mM glycine. Fixed cells were treated with permeabilisation solution (10 mM Tris-HCl, 10 mM NaCl, 0.2% IGEPAL CA-630). In situ digestion was performed using a cocktail of restriction enzymes (*Nla*III) to fragment chromatin while preserving crosslinked contacts. Digested DNA ends were ligated using T4 DNA ligase (New England Biolabs). Following ligation, crosslinks were reversed by overnight incubation at 65°C with proteinase K, and DNA was purified using phenol-chloroform extraction and ethanol precipitation. Purified Pore-C DNA was used to prepare Oxford Nanopore sequencing libraries using the Ligation Sequencing Kit (SQK-LSK114) according to the manufacturer's protocol. Libraries were loaded onto Prometh ION flow cells and sequenced on a PromethION24

instrument. Sequencing was performed for 72 hours per flow cell to achieve approximately 20× genome coverage per strain. Base calling was performed using Dorado v7.9.8 in Super-accurate mode.

5. Genome assembly pipeline

Genome assembly was performed using a multi-stage pipeline integrating PacBio HiFi reads for contig assembly and ONT Pore-C data for chromosome-level scaffolding. First, HiFi reads were assembled into contigs using NextDenovo v2.5.2 with default parameters (Hu *et al*., 2024). NextDenovo employs a correct-then-assemble strategy for long-read sequencing data to generate highly contiguous genome assemblies. Contig-level assemblies were polished using the original HiFi reads with Next Polish v1.4.1 to correct any residual errors (Hu *et al*., 2020). Polished contigs were then scaffolded to chromosome level using Pore-C data. Multi-way chromatin contacts were identified using pore-c-snakemake pipeline, which filters spurious contacts, normalizes contact frequencies, and generates contact matrices. Chromosome-scale scaffolding was performed using 3D-DNA v190716, which utilizes chromatin contact frequency information to detect mis-joins and to order and orient contigs into chromosome-scale scaffolds (Dudchenko *et al*., 2017). Manual curation of scaffolds was performed using Juicebox Assembly Tools v2.17 to correct mis-joins and resolve ambiguities based on visual inspection of Pore-C contact maps (Dudchenko *et al*., 2018). The curated assembly was finalized using the 3D-DNA post-review pipeline.

6. Quality assessment

Assembly quality was assessed using multiple complementary metrics. Contiguity was evaluated using standard statistics including total assembly size, number of scaffolds, N50 (the length of the shortest scaffold in the set of scaffolds that together represent at least 50% of the assembly), N90, maximum scaffold length, and average scaffold length. These metrics were calculated using QUAST v5.2.0 (Mikheenko *et al*., 2018). Completeness was assessed using BUSCO (Benchmarking Universal Single-Copy Orthologs) v5.8.3 with the apoidea_odb12 database, which contains 6,507 conserved single-copy orthologs expected to be present in Hymenoptera genomes. BUSCO classifies genes as complete (single-copy or duplicated), fragmented, or missing, providing a quantitative measure of assembly completeness and potential contamination (indicated by elevated duplication rates) (Seppey *et al*., 2019). Pore-C contact maps were generated using 3D-DNA v190716 and visualized using Juice box v2.17. Contact maps provide a visual representation of chromatin interaction frequencies across the genome, with strong diagonal blocks indicating contiguous chromosome-scale scaffolds and off-diagonal signals revealing structural features such as centromeres, telomeres, and large-scale rearrangements.

7. Gene ontology analysis

Functional annotation of the predicted gene models for all seven chromosome-level assemblies was performed using the OmicsBox platform (v3.2.9, Biobam, Spain), integrating sequence homology searches and domain signature identifications (Götz *et al*., 2008). First, protein-coding gene models were predicted using AUGUSTUS (v3.5.0) implemented in the OmicsBox under ab initio prediction mode, utilizing prediction models pre-trained specifically on *Apis*-specific gene structures. While direct transcriptomic evidence was not integrated during the prediction step, these ab initio models were subsequently validated and filtered through extensive sequence homology searches and domain consensus analyses. The resulting coding sequences (CDSs) were subsequently queried against the NCBI non-redundant (nr) protein database using BLASTx with an e-value significance threshold of 1×10^{-5} and a maximum of 20 database hits retained per sequence (Altschul *et al*., 1997; Sayers *et al*., 2025). Concurrently, functional domains, family signatures, and active sites were identified in silico by running InterProScan (v5.72-103.0) against multiple signature databases, including NCBIfam, SFLD, PANTHER, HAMAP, ProSite Profiles, CDD, PRINTS, Pfam, PIRSF, SUPERFAMILY, Gene3D, AntiFam, FunFam, and PIRSR (Jones *et al*., 2014). Gene Ontology (GO) terms representing three primary domains—Biological Process (BP), Molecular Function (MF), and Cellular Component (CC)—were assigned by merging the BLASTp mapping results and InterProScan domain annotations within the OmicsBox suite (Consortium, 2004). Enzyme Commission (EC) numbers were also mapped to sequences containing valid metabolic enzyme signatures. To perform high-resolution functional profiling, direct GO term counts (capturing

levels 3 to 7+) were extracted from the annotation outputs for each strain (e.g., go_count_bp.txt, go_count_mf.txt, and go_count_cc.txt) to examine granular functional differences. For comparative genomic analyses, GO level 2 distribution counts were normalized as percentages of the total GO-annotated genes within each strain to account for variations in gene-space prediction sizes. To determine if the overall functional landscape varied significantly among the seven honeybee breeding lines, a Chi-square (χ^2) test of independence was applied to the GO term contingency table. Subsequently, a two-tailed Welch's t-test (independent t-test with unequal variances) was performed to identify specific GO categories exhibiting statistically significant differences ($p < 0.05$) between *A. cerana* (n = 2) and *A. mellifera* (n = 5). Finally, a Z-score outlier analysis was conducted across the seven strains to identify unique, lineage-specific functional expansions ($Z > 2.0$). All statistical calculations, Welch's t-tests, and high-resolution multi-panel bar charts and heatmaps were executed in Python (v3.10) using scipy.stats, matplotlib, and seaborn libraries.

8. Phylogenetic analysis

To confirm the genetic identity and evolutionary relationships of the seven sequenced strains, we performed phylogenetic analysis based on complete mitochondrial genomes. The consensus contigs were assembled by NextDenovo (v2.5.2) from the selected HiFi reads corresponding to the bee reference mitochondrial genome, followed by concatenation using in-house perl script. The draft mitocondrial genome was completed to the circular form and polished using HiFi long reads by NextPolish (v1.4.1). The assembled mitochondrial genome was set to start from the same location with NC_014295.1 (*A. cerana*) or NC_051932.1 (*A. mellifera*). Mitochondrial genome sequences from the seven Korean strains were aligned with publicly available mitogenomes representing major *Apis* species and *A. mellifera* subspecies, including *A. m. ligustica*, *A. m. carnica*, *A. m. mellifera*, *A. m. anatoliaca*, and *A. cerana japonica*. To provide an evolutionary context, we included several *Apis* species as reference taxa (*A. florea*, *A. andreniformis*, *A. dorsata*, *A. laboriosa*). Multiple sequence alignment was performed using MAFFT v7.526 with the FFT-NS-2 algorithm. Phylogenetic trees were inferred using maximum likelihood (ML) in IQ-TREE v2.4.0 with 1,000 ultrafast bootstrap replicates to assess node support (Nguyen *et al*., 2015). The best-fit substitution model was selected based on the Bayesian Information Criterion (BIC). Trees were rooted using *Bombus ignitus* (GenBank accession: DQ870926) as the outgroup and visualized using FigTree v1.4.5 and MEGA11 (Katoh and Standley, 2013).

RESULTS

1. Geographic isolation and genetic background of conserved *Apis* breeding lines

Seven honeybee strains were successfully collected from two primary locations in Jeollabuk-do Province, South Korea: the Wido Isolated Breeding Station in Buan-gun and the National Institute of Agricultural Sciences in Jeonju. The five *A. mellifera* strains represented three major European subspecies: Italian bees (*A. m. ligustica*: Am-A, Am-C, Am-F), Carpatica/Carniolan-type bees (Am-D), and Carniolan bees (*A. m. carnica*: Am-V). The two *A. cerana* strains (Ac-R, Ac-X) were native Korean honeybees. All strains are registered in the Korean National Seed Bank and maintained under controlled breeding conditions to preserve genetic purity and prevent introgression from external populations. More than 50 individuals per strain were collected to ensure adequate sampling of within-colony genetic diversity.

2. High-depth long-read sequencing metrics and comparative performance of *de novo* genomic assemblers

PacBio HiFi sequencing on the Revio platform generated high-quality long-read data for all seven strains (Supplementary Table 1). The total number of HiFi reads per strain ranged from 777,008 (Am-C) to 1,116,667 (Am-V), corresponding to total yields of 10.35–12.45 Gbp. Read length N50 values ranged from 10,247 bp (Ac-R) to 17,038 bp (Am-C), with a median of approximately 13–14 kb for most strains. Median read quality scores were exceptionally high, ranging from Q36 to Q39, with 100% of reads exceeding Q20 (99% accuracy), 69–80% exceeding Q30 (99.9% accuracy), and 28–46% exceeding Q40 (99.99% accuracy). These quality metrics con-

Table 1. Comparative summary of genome assembly metrics between NextDenovo and Hifiasm pipelines

Assembler	Strain	Total assembly		N50 metrics		N90 metrics		Largest (bp)	Average (bp)
		Size (bp)	Contigs (#)	Length (bp)	Count (#)	Length (bp)	Count (#)		
Apis mellifera									
NextDenovo	Am-A	230,906,952	76	11,633,291	9	2,920,190	24	13,927,073	3,038,249
NextDenovo	Am-C	229,849,215	71	10,649,936	10	3,145,540	23	14,265,236	3,237,312
NextDenovo	Am-D	233,158,587	91	9,031,041	11	1,978,152	32	13,323,846	2,562,182
NextDenovo	Am-F	226,966,299	65	10,660,668	10	2,828,948	23	13,895,284	3,491,789
NextDenovo	Am-V	229,712,454	87	11,266,549	10	2,364,223	23	14,012,453	2,640,373
Apis cerana									
NextDenovo	Ac-R	216,051,137	130	6,004,893	14	1,298,854	44	11,793,743	1,661,931
NextDenovo	Ac-X	218,373,376	62	10,625,668	9	2,989,635	23	13,610,906	3,522,151
Apis mellifera									
Hifiasm	Am-A	278,976,762	486	14,231,502	8	627,794	27	28,075,939	574,026
Hifiasm	Am-C	274,979,850	363	12,305,727	10	922,540	30	16,848,399	757,520
Hifiasm	Am-D	267,626,722	309	12,684,975	9	1,527,761	24	26,748,018	866,106
Hifiasm	Am-F	264,318,499	287	13,990,627	8	3,102,612	18	28,059,269	920,970
Hifiasm	Am-V	267,612,102	352	12,633,761	9	1,034,819	23	18,586,869	760,262
Apis cerana									
Hifiasm	Ac-R	513,744,072	4,207	1,208,244	115	31,791	1,345	6,832,166	122,116
Hifiasm	Ac-X	508,355,635	1,875	1,782,925	69	101,609	496	9,235,821	271,123

firm the suitability of the HiFi data for high-fidelity genome assembly. The sequencing depth for each strain was estimated at approximately 45–55× coverage based on total yield and expected genome sizes (approximately 230 Mbp for *A. mellifera* and 220 Mbp for *A. cerana*). This exceeds the target coverage of 20× and provides substantial redundancy for accurate contig assembly and variant calling.

To evaluate assembly quality across the seven bee strains, we performed *de novo* assembly using two different long-read assemblers: NextDenovo and Hifiasm (Table 1). The resulting contig-level statistics showed distinct differences in total assembly size, contig numbers, and N50 lengths between the two methods. For the *A. mellifera* strains, NextDenovo generated assemblies with total sizes ranging from 226.96 Mb (Am-F) to 233.15 Mb (Am-D), comprised of 65 to 91 contigs. The contig N50 values for these assemblies ranged from 9.03 Mb (Am-D) to 11.63 Mb (Am-A). In comparison, Hifiasm produced larger assembly sizes for the same strains, ranging from 264.31 Mb (Am-F) to 278.97 Mb (Am-A), with higher contig counts ranging from 287 to 486, although it yielded contig N50 values between 12.31 Mb (Am-C) and 14.23 Mb (Am-A). In the *A. cerana* strains, the variance between the two assemblers was more pronounced. NextDenovo assembled the Ac-R and Ac-X genomes into 130 contigs (Size: 216.05 Mb, N50: 6.00 Mb) and 62 contigs (Size: 218.37 Mb, N50: 10.63 Mb), respectively. Conversely, Hifiasm assemblies for Ac-R and Ac-X resulted in substantially larger total sizes of 513.74 Mb and 508.36 Mb, with contig numbers increasing to 4,207 and 1,875, and contig N50 values of 1.21 Mb and 1.78 Mb, respectively. Based on the total assembly sizes and contig continuity metrics, the NextDenovo assemblies were selected as the representative genomes for all seven strains and utilized for subsequent downstream analyses.

3. Pore-C facilitated chromosomal scaffolding and species-specific scaffold length polymorphisms

The final chromosome-level assemblies captured a total length ranging from 207.68 Mb (Ac-R) to 223.22 Mb (Am-C) (Table 3). For all seven strains, the genomic seq-

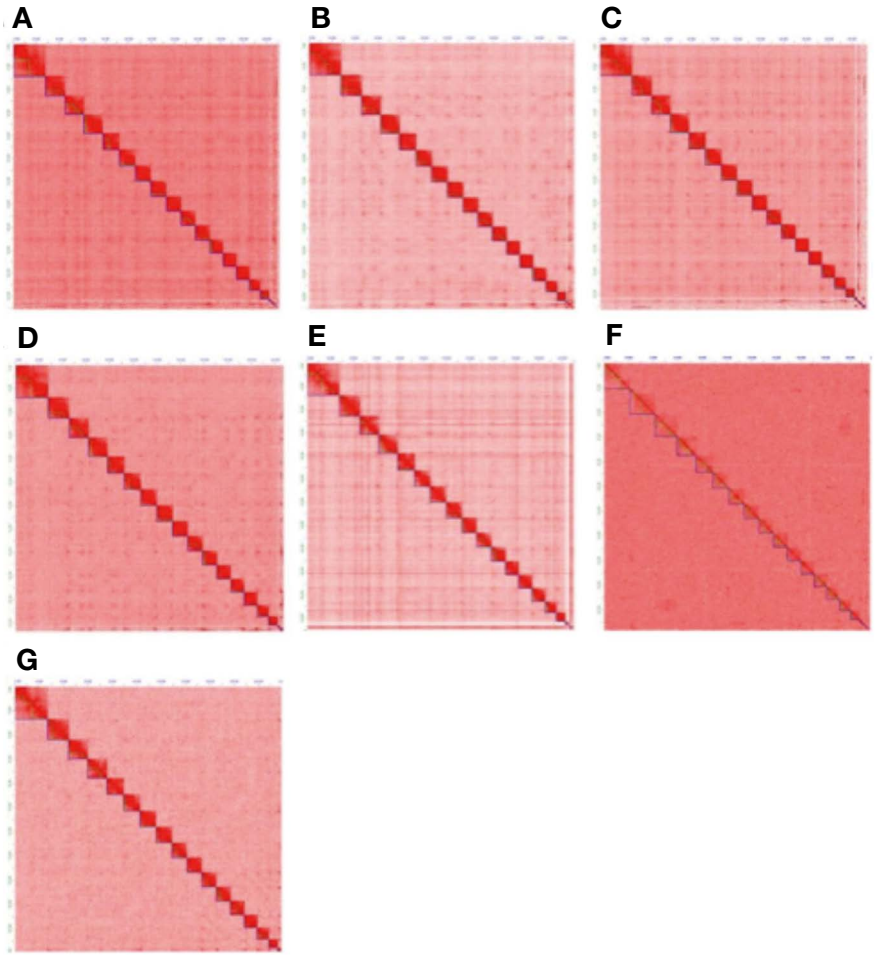


Fig. 1. Genome-wide Pore-C interaction heatmaps of seven reconstructed honeybee (*Apis*) strains. The individual contact matrices illustrate the chromatin interaction densities across the chromosome-level genome assemblies for five *A. mellifera* lineages—Am-A, Am-C, Am-D, Am-F, and Am-V—and two *A. cerana* lineages—Ac-R and Ac-X. The color intensity or pixel density reflects the frequency of localized physical genomic interactions, with the brightest diagonal signals representing contiguous genomic blocks. For all seven strains, the contact matrices display 16 sharply resolved, discrete interaction blocks along the diagonal axes. This configuration aligns precisely with the haploid chromosome number (n = 16) characteristic of the genus *Apis*, validating the structural completeness, high continuity, and robust anchoring of the *de novo* scaffolded pseudo-molecules without significant inter-chromosomal misassembles.

uences were successfully anchored into 16 major pseudo-molecules, which precisely corresponds to the haploid chromosome number (n = 16) of the genus *Apis* (Fig. 1, Supplementary Table 2). The genome-wide Pore-C interaction heatmaps demonstrated highly organized diagonal contact intensities with 16 distinct interaction blocks, validating the structural accuracy and continuity of the chromosomal scaffolding. The length distribution of the 16 chromosome-level scaffolds exhibited high consistency within each species, while displaying distinct species-specific profiles between *A. mellifera* and *A. cerana*. The five *A. mellifera* strains (Am-A, Am-C, Am-D, Am-F, and Am-V) showed nearly identical chromosomal length profiles across all lineages. In these strains, Scaffold_1 represented the largest chromosome, with lengths consistently clustered between 27.75 Mb and 27.79 Mb, whereas the smallest chromosome, Scaffold_16, ranged from 7.58 Mb to 7.63 Mb. The intermediate pseudo-molecules maintained highly uniform sizes among the five lineages, indicating strong structural conservation within the evaluated *A. mellifera* genomes. In contrast, the two *A. cerana* strains (Ac-R and Ac-X) presented intra-species variation in specific scaffold sizes compared to *A. mellifera*. In the Ac-R genome, Scaffold_1 measured 21.05 Mb, which was notably shorter than the 26.97 Mb observed in Ac-X. Conversely, Scaffold_2 and Scaffold_3 in Ac-R were

20.59 Mb and 18.01 Mb, respectively, both of which were larger than the corresponding scaffolds in Ac-X, which measured 17.17 Mb and 15.84 Mb. For the remaining pseudo-molecules from Scaffold_4 through Scaffold_16, both strains maintained comparable length distributions, with Scaffold_16 terminating at approximately 7.05 Mb for Ac-R and 7.26 Mb for Ac-X. These chromosomal statistics and the clear separation of the 16 contact matrices confirm that the assemblies achieved high-fidelity, chromosome-scale resolution suitable for subsequent downstream investigations.

4. Assessment of biological completeness and gene-space quality *via* ortholog verification

To evaluate the biological completeness and gene-space quality of the final scaffolded assemblies, a BUSCO analysis was performed using a core set of 6,507 orthologs (Table 2). The assessment demonstrated exceptionally high and uniform completeness across all seven evaluated bee genomes, with no significant structural or gene-content variation observed between the species or strains. For the *A. mellifera* strains, the total percentage of complete BUSCOs was remarkably consistent, ranging from 98.1% in Am-A, Am-C, Am-D, and Am-F to 98.2% in Am-V. Within these complete orthologs, single-copy genes accounted for 98.0% to 98.1% of the total dataset, while duplicated BUSCOs remained extremely low at a mere 0.1% (4 to 6 genes) for all five strains. The fraction of fragmented and missing BUSCOs was also tightly constrained, spanning only 0.5% (30 to 31 genes) and 1.4% to 1.5% (89 to 95 genes), respectively. A similarly high level of genomic completeness was observed in the *A. cerana* strains, with both Ac-R and Ac-X achieving a total complete BUSCO rate of 98.1%. Single-copy BUSCOs reached 98.0% for Ac-R and 97.9% for Ac-X, accompanied by a low duplication rate of 0.1% (7 and 8 genes, respectively). The fragmented genes constituted 0.5% to 0.6%, and the missing orthologs accounted for a minor fraction of 1.4% in both genomes.

5. Structural gene prediction fidelity and functional Gene Ontology (GO) profiling

Following the chromosome-level scaffolding, structural gene prediction and functional annotation were performed to evaluate the coding potential of the seven assembled genomes (Table 3). The total number of predicted protein-coding genes ranged from 6,860 (Ac-X) to 6,915 (Ac-R) in *A. cerana*, and from 7,261 (Am-F) to 8,751 (Am-D) in *A. mellifera*. To evaluate the accuracy of the *de novo* gene prediction models, the predicted gene sets were cross-referenced against multiple public functional databases, exhibiting exceptionally high homology rates. Across all seven strains, approximately 97% to 98% of the total predicted genes were successfully validated through alignment against the NCBI non-redundant (nr) database. Specifically, the nr mapping rates spanned from 97.20% (Ac-X) to 97.24% (Ac-R) for *A. cerana*, and from 96.94% (Am-A) to 97.58% (Am-V) for *A. mellifera*. Furthermore, protein domain analysis using InterProScan (IPS) successfully identified known functional domains in 65.71% to 66.22% of the *A. cerana* genes and 66.56% to 67.64% of the *A. mellifera* genes.

Table 2. BUSCO assessment results of genome assemblies for *Apis mellifera* and *Apis cerana* strains using the apoidea_odb12 lineage dataset

Assembly/Strain	Complete BUSCOs			Fragmented	Missing	Total BUSCOs
	Total complete	Single-copy	Duplicated			
Apis mellifera						
Am-A	6,386 (98.1%)	6,382 (98.1%)	4 (0.1%)	31 (0.5%)	90 (1.4%)	6,507 (100.0%)
Am-C	6,386 (98.1%)	6,380 (98.0%)	6 (0.1%)	31 (0.5%)	89 (1.4%)	6,507 (100.0%)
Am-D	6,383 (98.1%)	6,379 (98.0%)	4 (0.1%)	31 (0.5%)	93 (1.4%)	6,507 (100.0%)
Am-F	6,382 (98.1%)	6,378 (98.0%)	4 (0.1%)	30 (0.5%)	95 (1.5%)	6,507 (100.0%)
Am-V	6,388 (98.2%)	6,384 (98.1%)	4 (0.1%)	30 (0.5%)	89 (1.4%)	6,507 (100.0%)
Apis cerana						
Ac-R	6,381 (98.1%)	6,374 (98.0%)	7 (0.1%)	36 (0.6%)	90 (1.4%)	6,507 (100.0%)
Ac-X	6,381 (98.1%)	6,373 (97.9%)	8 (0.1%)	34 (0.5%)	92 (1.4%)	6,507 (100.0%)

Table 3. Summary of genome functional annotation metrics for *Apis mellifera* and *Apis cerana* strains

Species/Strain	Total predicted genes	NCBI non-redundant (nr)		InterProScan (IPS)		IPS and GO annotated		Merged GO annotated	
		Number of genes	Percentage (%)	Number of genes	Percentage (%)	Number of genes	Percentage (%)	Number of genes	Percentage (%)
Apis cerana									
Ac-R	6915	6782	98.08	6724	97.24	4544	65.71	6206	89.75
Ac-X	6860	6742	98.28	6668	97.20	4543	66.22	6205	90.45
Apis mellifera									
Am-A	8256	8073	97.78	8003	96.94	5527	66.95	7381	89.40
Am-C	7726	7578	98.08	7529	97.45	5215	67.50	6959	90.07
Am-D	8751	8603	98.31	8537	97.55	5919	67.64	7850	89.70
Am-F	7261	7132	98.22	7072	97.40	4833	66.56	6544	90.13
Am-V	8228	8100	98.44	8029	97.58	5553	67.49	7391	89.83

The integrated functional annotation pipeline culminated in successful Gene Ontology (GO) term assignment. Through the merging of IPS and blast-derived annotations, approximately 90% of the entire predicted gene complements were successfully assigned definitive GO terms. The merged GO annotation rates were highly uniform across all lineages, ranging from 89.75% to 90.45% in *A. cerana* and from 89.40% to 90.13% in *A. mellifera*. This high proportion of functionally characterized genes, combined with the comprehensive database alignment rates, demonstrates that the *de novo* gene prediction pipeline achieved substantial structural integrity without major sequence fragmentation or gene loss, confirming the high quality of the final genome annotations.

To further characterize the functional distribution of the annotated genes, the assigned GO terms were classified into Level 2 categories under three main ontologies: Biological Process (BP), Molecular Function (MF), and Cellular Component (CC) (Fig. 2). The overall distribution of GO terms exhibited a highly synchronized and consistent pattern across all seven evaluated strains. Within the Biological Process category, cellular process emerged as the most predominant term, encompassing approximately 76% of the GO-annotated genes in all lineages. This was followed by biological regulation and regulation of biological process, which consistently accounted for roughly 27–29% and 26–28% of the functional repertoire, respectively. Other well-represented BP subcategories included localization (~18%) and response to stimulus (~15%), demonstrating a highly uniform allocation of biological roles across both *A. mellifera* and *A. cerana* genomes. In the Molecular Function domain, binding and catalytic activity constituted the most dominant functional assignments. The binding category represented the largest fraction, with a stable proportion of approximately 83% across all seven strains. Catalytic activity was the second most abundant function, ranging between 39% and 47% of the annotated coding space. The remaining MF subcategories, such as transporter activity, transcription regulator activity, and ATP-dependent activity, presented identical hierarchical rankings and minor percentage variations among the lineages. For the Cellular Component ontology, the annotated genes were heavily concentrated in two primary structural categories: cellular anatomical structure, which represented over 80% of the annotated pool, and protein-containing complex, which comprised approximately 21–23% of the genes across all strains. The negligible divergence in GO term densities and distribution patterns across the entire dataset underlines that despite the genomic variance and lineage-specific divergence observed in structural statistics, the fundamental core proteome and functional machinery remain strictly conserved between *A. mellifera* and *A. cerana*.

6. Comparative functional enrichment and strain-specific toxicological and adaptive outliers

To capture functional variations within the conserved genomic architecture, a comparative enrichment analysis

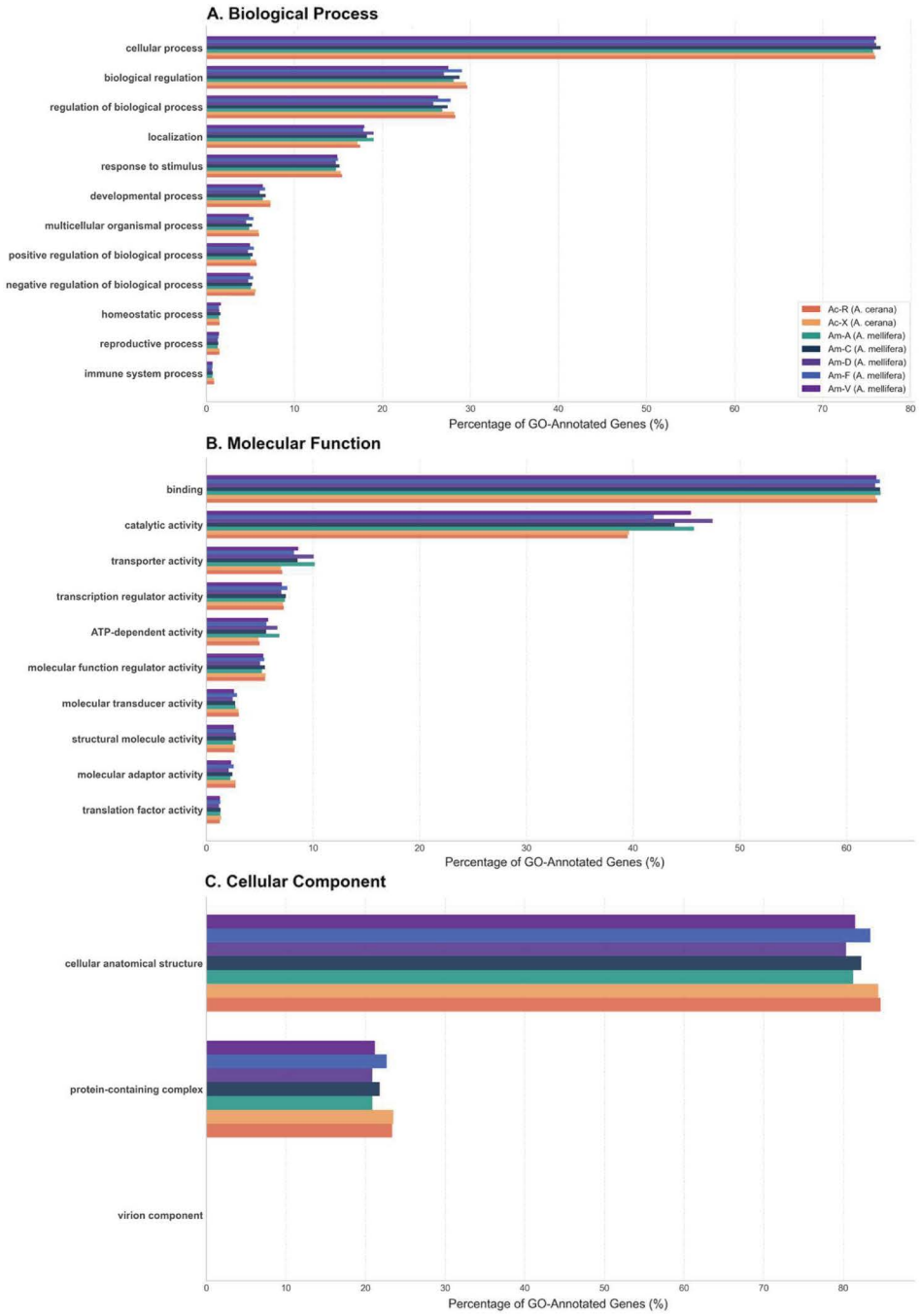


Fig. 2. Gene Ontology (GO) Level 2 functional classification of annotated protein-coding genes across seven *Apis* strains. The bar charts illustrate the distribution and percentage of functional categories annotated within the genomes of five *A. mellifera* strains (Am-A, Am-C, Am-D, Am-F, and Am-V) and two *A. cerana* strains (Ac-R and Ac-X). The classification is divided into three primary ontologies: (A) Biological Process, (B) Molecular Function, and (C) Cellular Component. The horizontal axis indicates the percentage of total GO-annotated genes within each individual lineage. Color-coded bars represent specific honeybee lineages, as defined in the species legend box. Comparative distribution displays a highly conserved and uniform functional landscape across all evaluated genomes, highlighting the stable retention of core physiological and metabolic machinery within the genus *Apis*.

of GO Level 2 terms was conducted using Fisher's exact test, followed by Benjamini-Hochberg False Discovery Rate (FDR) correction (adjusted $p<0.05$), based on the pooled counts of annotated genes across all genomes

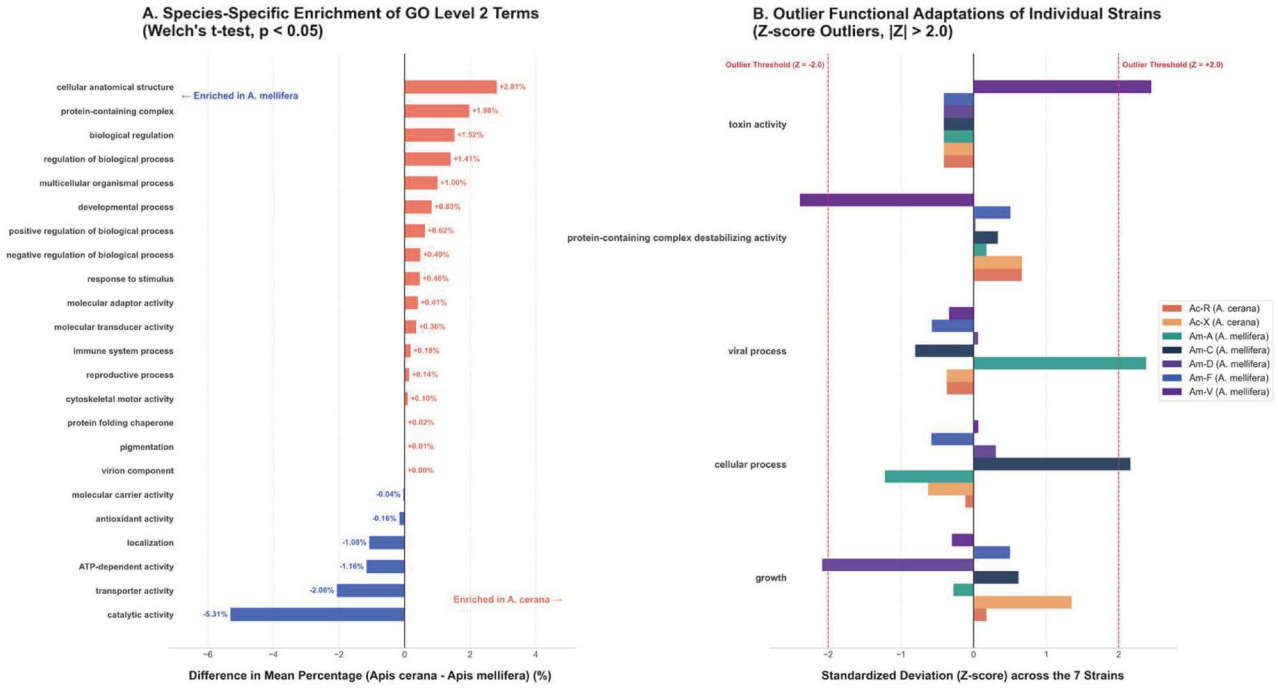


Fig. 3. Comparative enrichment and strain-specific outlier analyses of GO Level 2 terms among *Apis* lineages. (A) Species-specific functional enrichment under Fisher's exact test (FDR-adjusted $p<0.05$). The horizontal axis displays the difference in the relative percentage of annotated genes between *A. cerana* and *A. mellifera* (MeanAc - MeanAm). Red bars extending to the right represent functional categories significantly enriched in *A. cerana*, while blue bars extending to the left indicate categories significantly overrepresented in *A. mellifera*. Statistical significance levels (*p*-values) and mean difference values are explicitly labeled on individual bars. (B) Outlier functional adaptations of individual strains based on Z-score profiling across the seven genomes. The horizontal axis represents the standardized deviation (Z-score) of GO term proportions for each strain against the group mean. Vertical red dashed lines denote the statistical outlier thresholds ($|Z|=2.0$). Color-coded bars mirror the specific honeybee lineages indicated in the color key. Prominent lineage-specific signatures are highlighted, including the acute elevation of toxin activity in the Am-V strain ($Z=2.45$), revealing highly localized and specialized evolutionary selection pressures acting upon individual genomic landscapes.

(Fig. 3A, Supplementary Table 3). *A. cerana* genomes exhibited significant enrichment in pathways associated with cellular anatomical structure (+2.88%, adjusted $p<0.001$) and protein-containing complexes (+2.02%, adjusted $p<0.001$). This structural profile was further associated with enrichment in biological regulation (+1.57%, adjusted $p=0.005$), developmental processes (+0.84%, adjusted $p=0.007$), and regulation of biological processes (+1.45%, adjusted $p=0.009$). Conversely, *A. mellifera* genomes were significantly enriched in categories driving metabolic efficiency and transport. The most pronounced divergence was observed in catalytic activity, which showed a significant difference of 5.42% relative to *A. cerana* (adjusted $p<0.001$). Additionally, transporter activity was significantly enriched in *A. mellifera* (2.10%, adjusted $p<0.001$), along with ATP-dependent activity (1.18%, adjusted $p<0.001$) and localization (1.10%, adjusted $p=0.024$).

Beyond the species-level evolutionary divergence, a standardized Z-score analysis ($|Z|>2.0$) was deployed to screen for strain-specific functional outliers across the seven evaluated lineages (Fig. 3B). This filtration identified localized genomic configurations unique to individual bee strains that deviated significantly from the cross-lineage baseline. A distinct functional outlier signature was captured in the *A. mellifera* Carniolan-derived strain, Am-V, which displayed a pronounced statistical overrepresentation specifically within the toxin activity category ($Z=2.45$). Furthermore, the Z-score profiling revealed positive deviations shared between the two Carniolan lineages, Am-D and Am-V, in specific somatic and metabolic categories. In the Z-score distribution of these two strains, this profile was characterized by the statistically significant overrepresentation of crucial metabolic and

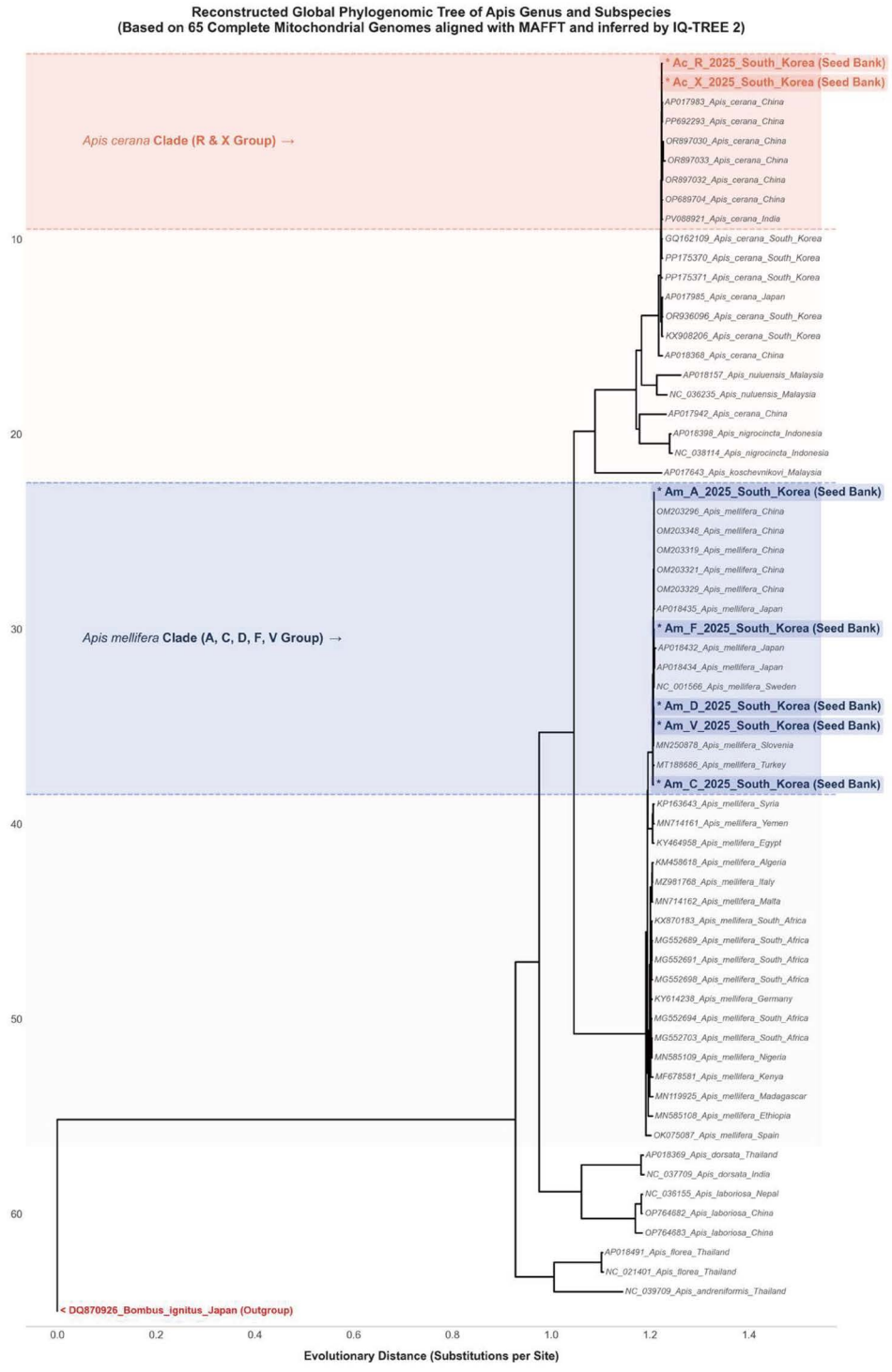


Fig. 4. Reconstructed global phylogenomic tree of the genus *Apis* and its subspecies based on complete mitogenomes. The Maximum Likelihood (ML) phylogenetic topology was inferred using IQ-TREE 2 based on the alignment of 65 complete mitochondrial genomes *via* MAFFT. *Bombus ignitus* (GenBank accession: DQ870926) was utilized as the evolutionary outgroup to root the tree. The horizontal scale bar at the bottom indicates the evolutionary distance, measured in substitutions per site. The red shaded box highlights the *A. cerana* continental clade (y = 1.0 to 9.0 region), positioning the two Korean seed bank strains (Ac-R and Ac-X, marked with asterisks) strictly within the mainland East Asian lineage alongside Chinese and Indian accessions. The blue shaded box outlines the *A. mellifera* evolutionary lineage (y = 23.0 to 38.0 region), mapping the five Korean seed bank strains (Am-A, Am-C, Am-D, Am-F, and Am-V, marked with asterisks) exclusively into the European C-lineage group.

xenobiotic detoxification gene families, including cytochrome P450 monooxygenases. The quantitative convergence of these amplified detoxification pathways and metabolic homeostasis categories was uniquely localized to the Am-D and Am-V datasets compared to the other three western honeybee lineages.

7. Mitochondrial genome characterization and phylogenetic placement

To determine the maternal lineages and evolutionary origins of the seven honeybee strains, we successfully reconstructed their complete, gapless circular mitochondrial genomes utilizing the high-fidelity long reads from the PacBio Revio platform. The assembled mitogenomes exhibited highly conserved structural architectures and species-specific length distributions (Supplementary Table 4). The two *A. cerana* strains (Ac-R and Ac-X) possessed concise genomes ranging from 16,224 bp to 16,247 bp. In comparison, the five *A. mellifera* strains (Am-A, Am-C, Am-D, Am-F, and Am-V) displayed slightly larger but exceptionally uniform sizes, spanning from 16,515 bp to 16,576 bp. All seven mitogenomes exhibited an extreme AT-bias, with the total AT content ranging between 84.29% and 85.09%.

To evaluate the phylogenetic positions of these lineages, a comprehensive 65-taxon Maximum Likelihood (ML) phylogenetic tree was constructed based on the complete mitogenome sequences (Fig. 4, Supplementary Table 5). The phylogenetic topology revealed a clear genetic dichotomy that separated the eastern and western honeybee lineages into distinct, well-supported geographic clades. The two Korean *A. cerana* genebank strains (Ac-R and Ac-X) clustered together with absolute bootstrap support within the mainland East Asian continental clade (y = 1.0 to 9.0 region), positioning adjacent to reference accessions from China and India. Notably, this mainland continental clade was phylogenetically distinct from the Japanese *A. cerana* lineage (AP017985), as well as from other wild genetic resources native to Southeast and Southwest Asia. Conversely, the five *A. mellifera* strains maintained by the national genebank (Am-A, Am-C, Am-D, Am-F, and Am-V) were unequivocally assigned to the European C-lineage (y = 23.0 to 38.0 region), which encompasses major Mediterranean and Eastern European subspecies such as *A. m. carnica* and *A. m. ligustica*. Although all five strains anchored firmly within the C-lineage, they exhibited subtle branch-length variations and distinct sub-clade bifurcations, revealing a meaningful degree of intra-lineage genetic diversity. Particularly, the Am-C strain clustered closely with Eastern Mediterranean gene pools, including the *A. m. anatoliaca* subspecies, in addition to the typical Western European references.

To overcome the maternal-inheritance limitation of mitochondrial phylogeny and confirm nuclear breeding-line integrity, we also performed a whole-genome nuclear distance comparison using Mash (k = 21, s = 10,000) (Supplementary Fig. 1). The nuclear distance matrix revealed a clear genetic divergence between *A. cerana* and *A. mellifera* strains (average distance of 0.058, equivalent to 94.2% Average Nucleotide Identity [ANI]). Within species, the nuclear pairwise distances were tightly constrained (~0.005 distance, 99.5% ANI), with the two Korean *A. cerana* seed bank strains (Ac-R and Ac-X) clustering closely together. This dual phylogenomic validation confirms the genetic purity, breed integrity, and maternal-paternal lineage stability of all seven national seed bank repositories.

DISCUSSION

In this study, we established a definitive, chromosome-scale genomic framework for seven economically and ecologically vital honeybee strains maintained by the national seed bank of South Korea. Utilizing the high-fidelity long reads of the PacBio Revio platform coupled with Pore-C chromosomal conformation capture, we overcame the long-standing computational bottlenecks associated with assembling highly heterozygous, AT-rich (> 84%), and repeat-dense aculeate genomes (Elsik *et al*., 2014; Wallberg *et al*., 2019). While alternative assembly strategies like Hifiasm suffered from severe pseudo-haplotypic inflation and sequence fragmentation—particularly in the *A. cerana* lineages—our optimized Next Denovo pipeline successfully resolved the genomic architectures into the exact haploid chromosome count (n = 16) with exceptional continuity (contig N50 up to 11.63 Mb). Furthermore, this high structural accuracy directly translated into superior downstream gene predictions, yield-

ing complete functional annotations for approximately 90% of the entire coding space without structural fragmentation or gene-space loss. This technical leap, combined with the successful gapless reconstruction of complete circular mitogenomes, marks the first time that the core genetic repositories of both *A. mellifera* and *A. cerana* in Korea have been mapped with such exhaustive completeness.

The innovation of these datasets extends far beyond technical precision, serving as a foundational digital infrastructure for precision breeding and adaptive evolutionary genomics. Historically, honeybee genomic initiatives have been constrained by fragmented reference templates, which inherently introduce systemic biases during structural variant calling and functional pathway profiling (Wallberg *et al*., 2014; Howe *et al*., 2021). By delivering highly curated, line-specific chromosome-level assemblies, this study provides an un-biased genomic ledger to trace fine-scale functional adaptations and evolutionary divergence between eastern and western honeybee species. As apiculture faces accelerating threats from climate change, habitat loss, and emerging pathogens, these high-fidelity genomes offer the precise molecular resolution needed to dissect complex physiological networks, such as temperature-induced metabolic shifts and sensory-mediated pest resistance (Potts *et al*., 2010; Goulson *et al*., 2015). Consequently, these genomic data establish an indispensable reference framework that underpins future molecular breeding platforms, facilitating the strategic conservation and sustainable utilization of regional honeybee genetic resources.

1. Chromosome-scale length variations and implications for population divergence

A compelling structural finding in our genomic survey was the intra-species scaffold length variance observed between the two *A. cerana* lineages, Ac-R and Ac-X. Specifically, scaffold_1 in Ac-R was found to be approximately 5.92 Mb shorter than its homologue in Ac-X, whereas scaffold_2 and scaffold_3 in Ac-R exhibited reciprocal size increases of 3.42 Mb and 2.16 Mb, respectively. It is critical to underscore that while formal structural variant (SV) discovery pipelines—such as long-read split-mapping or optical mapping—were not deployed in the current study, these macro-level length disparities captured *via* Pore-C scaffolding provide indirect yet tangible genomic signatures of large-scale chromosomal rearrangements. In hymenopteran genomics, such pronounced chromosomal length discrepancies among closely related allopatric or sympatric populations typically point to massive structural modifications, most notably chromosomal inversions, translocations, or the localized lineage-specific accumulation of transposable elements (TEs) (Kapusta *et al*., 2017; Mérot *et al*., 2020). Large chromosomal inversions, for instance, frequently act as evolutionary "supergenes" by physically suppressing meiotic recombination between divergent lineages (Wellenreuther and Bernatchez, 2018). This recombination barrier locks adaptive allele complexes together, allowing populations to maintain distinct ecological phenotypes—such as specialized overwintering strategies, foraging behaviors, or predatory defense mechanisms—despite potential gene flow (Wallberg *et al*., 2017).

Given that Ac-R and Ac-X represent distinct indigenous conservation lines within the Korean peninsula, the identified polymorphic regions on scaffolds 1, 2, and 3 stand out as primary candidates for maintaining unique agronomic traits. To transition from these macro-structural observations to definitive mechanistic insights, a comprehensive, high-resolution SV profiling initiative utilizing long-read graph-pangenome mapping remains an essential next step (Koren *et al*., 2018; Garg *et al*., 2021). Characterizing the exact breakpoints and sequence content of these polymorphic blocks will be paramount to unlocking the genetic basis of fine-scale adaptive divergence in Korean *A. cerana*, directly empowering marker-assisted precision breeding and regional ecotype preservation (Meuwissen *et al*., 2001; Brascamp *et al*., 2018).

2. Species- and strain-specific functional divergence and phenotypic correlates

The comparative statistical framework identified targeted functional shifts that suggest potential genomic underpinnings for long-observed phenotypic, behavioral, and agronomic disparities between eastern and western honeybees, although direct links between broad GO Level 2 terms and complex phenotypes require cautious interpretation. The enrichment of biological regulation and regulation of biological process in *A. cerana* may reflect its complex regulatory adaptations. Historically, *A. cer-*

ana has demonstrated remarkable ecological resilience against devastating endemic parasites and pathogens, such as *Varroa* mites and the Sacbrood virus, largely driven by highly specialized hygienic behaviors including intensive grooming and targeted removal of infected brood (Boecking and Genersch, 2008; Oxley *et al*., 2010). While broad GO categories such as response to stimulus or immune system process do not show significant species-wide expansion in overall gene counts under strict FDR correction, the specific gene families nested within these categories—including chemosensory receptors and specific immune signaling cascades detailed in subsequent sections—provide the precise molecular repertoire supporting these rapid, systemic responses. Furthermore, the robust representation of cellular and regulatory pathways provides a plausible genetic background that may facilitate the acute behavioral vigilance of *A. cerana*, potentially modulating rapid colony-level alarm signaling and collective defensive tactics—such as heat-balling behaviors—when encountering aggressive native predators like *Vespa mandarinia* (Tan *et al*., 2012).

Conversely, the genomic architecture of *A. mellifera* appears evolutionarily tailored toward maximized metabolic output and resource acquisition. The pronounced enrichment in catalytic activity (adjusted $p < 0.001$, −5.42%) and transporter activity (adjusted $p < 0.001$, −2.10%) provides a compelling genetic framework that may underlie the high foraging efficiency and commercial productivity that define the western honeybee (Wallberg *et al*., 2014; Uzunov *et al*., 2017). These metabolic expansions suggest a potentially intensified enzymatic machinery dedicated to rapid carbohydrate processing, ATP-driven flight-energy generation, and systemic nutrient transport (Foret *et al*., 2012; Kapheim *et al*., 2015). This amplified metabolic throughput directly translates into the extended foraging ranges, prolonged flight capacities, and superior honey-storing behaviors observed in agricultural settings, reflecting the plausible historical selection pressures that favored high-yield traits in *A. mellifera* (Harpur *et al*., 2012).

At the intra-species level, the standardized Z-score filtering successfully exposed localized, hyper-specialized adaptations within individual commercial lines, most notably in the Carniolan-derived genotypes. The sharp statistical deviation in the toxin activity category observed exclusively in the Am-V strain ($Z = 2.45$) marks it as an evolutionary outlier. Detailed sequence analysis of the genes driving this category in Am-V revealed a combination of host defensive peptides and gut symbiont sequences. Specifically, we identified omega-conotoxin-like protein 1 (also known as inhibitor cysteine knot peptide), a eukaryotic disulfide-rich peptide known to be involved in arthropod defense and venom. However, the outlier status was also driven by the presence of two co-assembled hypothetical proteins from the honeybee gut symbiont *Snodgrassella alvi* (Supplementary Table 6). This indicates that the apparent statistical enrichment of toxin activity in Am-V is a composite signature, reflecting both host defense genes and co-assembled symbiotic microbiome elements rather than a massive host-specific duplication of venom components. Furthermore, the distinct functional configurations shared by the Carniolan sister strains, Am-D and Am-V, provide a molecular link to their renowned overwintering capacity and cold hardiness (Büchler *et al*., 2014). Cross-strain copy number comparisons of the Cytochrome P450 (CYP) gene family (Supplementary Table 7) revealed highly conserved profiles across all seven strains, with 18 genes in *A. cerana* and 19 to 21 genes in *A. mellifera* (Am-A: 19, Am-C: 21, Am-D: 20, Am-F: 21, Am-V: 21), indicating that no gene family expansion occurred in the Carniolan strains. Thus, their cold-adaptation capabilities are likely regulated through transcriptional modulation of these conserved detoxifying pathways and metabolic homeostatic systems, rather than genomic copy number expansion. This conserved enzymatic buffer, combined with lineage-specific expression dynamics, equips these specific lineages to effectively mitigate low-temperature stress and cellular toxicity (Guarna *et al*., 2017), contributing to their survival rates during harsh wintering periods.

3. Phylogeographic origins and genetic identity of national repositories

The gapless reconstruction of complete circular mitogenomes and the subsequent 65-taxon phylogenomic mapping provide unequivocal maternal evolutionary contexts for the honeybee genetic resources conserved in South Korea. For the indigenous eastern honeybee strains, Ac-R and Ac-X, their robust positioning within the mainland East Asian continental clade ($y = 1.0$ to 9.0)

settles a critical question regarding their phylogeographic ancestry (Radloff *et al*., 2010). By exhibiting a clear genetic divergence from the Japanese *A. cerana* lineage (AP 017985) and wild island ecotypes of Southeast Asia, our findings demonstrate that the Korean native bee repository shares a direct evolutionary trajectory with the mainland Eurasian continental stock (Takahashi *et al*., 2016; Nunn *et al*., 2022). This biogeographic assignment is highly congruent with multi-locus nuclear screening demonstrating that South Korean *A. cerana* stocks form a highly integrated, uniform temperate cluster alongside Russian Far East populations, sharply distinguished from tropical Southeast Asian ecotypes (Ilyasov *et al*., 2022; Kaskinova *et al*., 2022). Critically, this regional gene pool bears the molecular signatures of severe evolutionary bottlenecks and intense purifying selection driven by historical Sacbrood virus epidemics that previously eliminated up to 95% of the domestic colonies, validating our conserved seed bank strains as indispensable, highly screened genetic reservoirs for regional ecotype preservation. This continental genomic signature implies a shared ancestral adaptation to severe temperate seasonal fluctuations, serving as a highly valuable genetic baseline for the preservation of pure domestic ecotypes and the development of robust, climate-resilient local breeding lines (Chen *et al*., 2016).

Parallel to this, the phylogenomic clustering of the five western honeybee strains (Am-A, Am-C, Am-D, Am-F, and Am-V) exclusively within the European C-lineage (y = 23.0 to 38.0) confirms the structural retention of their Mediterranean and Eastern European maternal ancestry (De la Rúa *et al*., 2009; Boardman *et al*., 2019). More importantly, the identification of micro-clade bifurcations and subtle branch-length variations among these five lines reveals that they do not constitute a homogenous, bottlenecked population. The close phylogenetic affinity of the Am-C strain toward Eastern Mediterranean gene pools, such as *A. m. anatoliaca*, alongside typical *A. m. carnica* or *A. m. ligustica* frameworks, serves as a molecular footprint of historical genetic admixture. Rather than representing isolated, singular western imports, these national conservation lines function as dynamic genetic reservoirs (Harpur *et al*., 2012, 2015). Over decades of localized maintenance, artificial selection, and selective breeding within the Korean peninsula, these strains have integrated diverse elite European genetic components. This hybrid-driven genomic plasticity likely accelerated their regional acclimatization, resulting in a unique, domestic-adapted gene pool tailored to the specific environmental and apicultural demands of South Korea.

CONCLUSION

Taken together, this study represents a major milestone in apicultural genomics by delivering the first comprehensive, chromosome-level reference atlas for both *A. mellifera* and *A. cerana* conservation lines in Korea. By overcoming historical assembly limitations through an optimized long-read and Pore-C integration pipeline, we successfully mapped their structural, functional, and mitogenomic landscapes with unprecedented fidelity. The distinct species-specific GO enrichments and the macro-structural variances identified across these lineages underscore the profound genetic potential available for strategic stock improvement. Looking forward, these high-resolution templates will serve as the indispensable digital infrastructure for next-generation precision breeding. They will enable the high-throughput development of molecular markers, the precise tracing of adaptive sensory-receptor repertoires (Kohno and Kubo, 2019), and the comprehensive mapping of metabolic stress pathways under shifting global climate regimes, ultimately ensuring the long-term sustainability and security of regional honeybee ecosystems (Allendorf *et al*., 2010; Spötter *et al*., 2012).

ACKNOWLEDGEMENTS

We sincerely express our gratitude to Soyeon Kang, Sanghee Um, and Hwayong An from the National Instrumentation Center for Environmental Management (NICEM) at Seoul National University for their dedicated technical support in high-molecular-weight DNA extraction, sequencing library preparation, and ONT Pore-C proximity-ligation sequencing assays. We also thank the staff of the Wido Isolated Breeding Station (Buan-gun, Jeollabuk-do) and the National Institute of Agricultural Sciences (Jeonju, Jeollabuk-do) for assist-

ance with sample collection and maintenance of honeybee breeding lines. This research was funded by "The Cooperative Research Program for Agriculture Science & Technology Development" (Project No. RS-2025-02214782) from the Rural Development Administration, Republic of Korea. We declare no competing interests.

LITERATURE CITED

Allendorf, F. W., P. A. Hohenlohe and G. Luikart. 2010. Genomics and the future of conservation genetics. Nat. Rev. Genet. 11(10): 697-709.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25(17): 3389-3402.

Boardman, L., A. Eimanifar, R. T. Kimball, E. L. Braun, S. Fuchs, B. Grünewald and J. D. Ellis. 2019. The mitochondrial genome of the Carniolan honey bee, *Apis mellifera carnica* (Insecta: Hymenoptera: Apidae). Mitochondrial DNA Part B 4(2): 3288-3290.

Boecking, O. and E. Genersch. 2008. Varroosis - the ongoing crisis in bee keeping. J. Consum. Prot. Food Saf. 3(2): 221-228.

Brascamp, P. and P. Bijma. 2018. Methods to estimate genetic parameters and breeding values for honey bees. Genet. Sel. Evol. 50(1): 45.

Büchler, R., C. Costa, F. Hatjina, S. Andonov, M. D. Meixner, Y. L. Conte, A. Uzunov, S. Berg, M. Bienkowska and M. Bouga. 2014. The influence of genetic origin and its interaction with environmental effects on the survival of *Apis mellifera* L. colonies in Europe. J. Apic. Res. 53(2): 205-214.

Chapman, N. C., B. A. Harpur, J. Lim, T. E. Rinderer, M. H. Allsopp, A. Zayed and B. P. Oldroyd. 2015. A SNP test to identify Africanized honeybees via proportion of 'African' ancestry. Mol. Ecol. Resour. 15(6): 1346-1355.

Chen, C., Z. Liu, Q. Pan, X. Chen, H. Wang, H. Guo, S. Liu, H. Lu, S. Tian and R. Li. 2016. Genomic analyses reveal demographic history and temperate adaptation of the newly discovered honey bee subspecies *Apis mellifera sinisxinyuan* n. ssp. Mol. Biol. Evol. 33(5): 1337-1348.

Cheng, H., G. T. Concepcion, X. Feng, H. Zhang and H. Li. 2021. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. Nat. Methods 18(2): 170-175.

Consortium, G. O. 2004. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. 32(suppl_1): D258-D261.

Crane, E. 1999. The world history of beekeeping and honey hunting. Routledge, New York. 704 pp.

De la Rúa, P., R. Jaffé, R. Dall'Olio, I. Muñoz and J. Serrano. 2009. Biodiversity, conservation and current threats to European honeybees. Apidologie 40(3): 263-284.

Dudchenko, O., S. S. Batra, A. D. Omer, S. K. Nyquist, M. Hoeger, N. C. Durand, M. S. Shamim, I. Machol, E. S. Lander and A. P. Aiden. 2017. *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science 356(6333): 92-95.

Dudchenko, O., M. S. Shamim, S. S. Batra, N. C. Durand, N. T. Musial, R. Mostofa, M. Pham, B. Glenn St Hilaire, W. Yao and E. Stamenova. 2018. The Juicebox Assembly Tools module facilitates *de novo* assembly of mammalian genomes with chromosome-length scaffolds for under $1000. BioRxiv 254797.

Elsik, C. G., K. C. Worley, A. K. Bennett, M. Beye, F. Camara, C. P. Childers, D. C. de Graaf, G. Debyser, J. Deng, B. Devreese, E. Elhaik, J. D. Evans, L. J. Foster, D. Graur, R. Guigo, K. J. Hoff, M. E. Holder, M. E. Hudson, G. J. Hunt, H. Jiang, V. Joshi, R. S. Khetani, P. Kosarev, C. L. Kovar, J. Ma, R. Maleszka, R. F. A. Moritz, M. C. Munoz-Torres, T. D. Murphy, D. M. Muzny, I. F. Newsham, J. T. Reese, H. M. Robertson, G. E. Robinson, O. Rueppell, V. Solovyev, M. Stanke, E. Stolle, J. M. Tsuruda, M. Van Vaerenbergh, R. M. Waterhouse, D. B. Weaver, C. W. Whitfield, Y. Wu, E. M. Zdobnov, L. Zhang, D. Zhu and R. A. Gibbs. 2014. Finding the missing honey bee genes: lessons learned from a genome upgrade. BMC Genom. 15: 86.

Foret, S., R. Kucharski, M. Pellegrini, S. Feng, S. E. Jacobsen, G. E. Robinson and R. Maleszka. 2012. DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. Proc. Natl. Acad. Sci. 109(13): 4968-4973.

Frankham, R., J. D. Ballou and D. A. Briscoe. 2010. Introduction to conservation genetics. Cambridge University Press, Cambridge. 642 pp.

Frunze, O., P. N. Akongte, D. Kim, E.-J. Kang, K. Kim, B.-S. Park and Y.-S. Choi. 2022. Remote honey bee breeding center in the Wido Island, the Republic of Korea. J. Apic. 37(2): 175-183.

Garg, S., A. Fungtammasan, A. Carroll, M. Chou, A. Schmitt, X. Zhou, S. Mac, P. Peluso, E. Hatas and J. Ghurye. 2021. Chromosome-scale, haplotype-resolved assembly of human genomes. Nat. Biotechnol. 39(3): 309-312.

Goel, M., H. Sun, W.-B. Jiao and K. Schneeberger. 2019. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biol. 20(1): 277.

Götz, S., J. M. García-Gómez, J. Terol, T. D. Williams, S. H. Nagaraj, M. J. Nueda, M. Robles, M. Talón, J. Dopazo and A. Conesa. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36(10): 3420-3435.

Goulson, D., E. Nicholls, C. Botías and E. L. Rotheray. 2015. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. Science 347(6229): 1255957.

Guarna, M. M., S. E. Hoover, E. Huxter, H. Higo, K.-M. Moon, D. Domanski, M. E. Bixby, A. P. Melathopoulos, A. Ibrahim and M. Peirson. 2017. Peptide biomarkers used for the selective breeding of a complex polygenic trait in honey bees. Sci. Rep. 7(1): 8381.

Harpur, B., N. Chapman, L. Krimus, P. Maciukiewicz, V. Sandhu, K. Sood, J. Lim, T. Rinderer, M. Allsopp and B. Oldroyd. 2015. Assessing patterns of admixture and ancestry in Canadian honey bees. Insectes Soc. 62(4): 479-489.

Harpur, B. A., S. Minaei, C. F. Kent and A. Zayed. 2012. Management increases genetic diversity of honey bees via admixture. Mol. Ecol. 21(18): 4414-4421.

Howe, K., W. Chow, J. Collins, S. Pelan, D.-L. Pointon, Y. Sims, J. Torrance, A. Tracey and J. Wood. 2021. Significantly improving the quality of genome assemblies through curation. GigaScience 10(1): giaa153.

Hu, J., J. Fan, Z. Sun and S. Liu. 2020. NextPolish: a fast and efficient genome polishing tool for long-read assembly. Bioinformatics 36(7): 2253-2255.

Hu, J., Z. Wang, Z. Sun, B. Hu, A. O. Ayoola, F. Liang, J. Li, J. R. Sandoval, D. N. Cooper and K. Ye. 2024. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. Genome Biol. 25(1): 107.

Ilyasov, R. A., S. Rašić, J. Takahashi, V. N. Danilenko, M. Y. Proshchalykin, A. S. Lelej, V. N. Sattarov, P. H. Thai, R. Raffiudin and H. W. Kwon. 2022. Genetic relationships and signatures of adaptation to the climatic conditions in populations of *Apis cerana* based on the polymorphism of the gene vitellogenin. Insects 13(11): 1053.

Jones, P., D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell and G. Nuka. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9): 1236-1240.

Jung, C. and S. G. Lee. 2018. Beekeeping in Korea: Past, present, and future challenges. J. Apic. 33(2): 127-135.

Kapheim, K. M., H. Pan, C. Li, S. L. Salzberg, D. Puiu, T. Magoc, H. M. Robertson, M. E. Hudson, A. Venkat and B. J. Fischman. 2015. Genomic signatures of evolutionary transitions from solitary to group living. Science 348(6239): 1139-1143.

Kapusta, A., A. Suh and C. Feschotte. 2017. Dynamics of genome size evolution in birds and mammals. Proc. Natl. Acad. Sci. 114(8): E1460-E1469.

Kaskinova, M., L. Gaifullina, R. Ilyasov, A. Lelej, H. W. Kwon, P. H. Thai and E. Saltykova. 2022. Genetic structure of *Apis cerana* populations from South korea, Vietnam and the Russian Far East based on microsatellite and mitochondrial DNA polymorphism. Insects 13(12): 1174.

Katoh, K. and D. M. Standley. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol. Biol. Evol. 30(4): 772-780.

Klein, A.-M., B. E. Vaissière, J. H. Cane, I. Steffan-Dewenter, S. A. Cunningham, C. Kremen and T. Tscharntke. 2007. Importance of pollinators in changing landscapes for world crops. Proc. Royal Soc.y B: Biol. Sci. 274(1608): 303-313.

Kohno, H. and T. Kubo. 2019. Genetics in the honey bee: achievements and prospects toward the functional analysis of molecular and neural mechanisms underlying social behaviors. Insects 10(10): 348.

Koren, S., A. Rhie, B. P. Walenz, A. T. Dilthey, D. M. Bickhart, S. B. Kingan, S. Hiendleder, J. L. Williams, T. P. Smith and A. M. Phillippy. 2018. De novo assembly of haplotype-resolved genomes with trio binning. Nat. Biotechnol. 36(12): 1174-1182.

Lee, J.-H., B.-J. Kim, G. Han, O. Frunze, G. Nah and H. W. Kwon. 2025. Chromosome level de Novo hybrid assembly of Asian honeybee, *Apis cerana Koreana*. Sci. Rep. 15(1): 26912.

Lieberman-Aiden, E., N. L. Van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo and M. O. Dorschner. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326(5950): 289-293.

Meixner, M. D. 2010. A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. J. Inverteb. Pathol. 103: S80-S95.

Meixner, M. D., C. Costa, P. Kryger, F. Hatjina, M. Bouga, E. Ivanova and R. Büchler. 2010. Conserving diversity and vitality for honey bee breeding. J. Apic. Res. 49(1): 85-92.

Mérot, C., R. A. Oomen, A. Tigano and M. Wellenreuther. 2020. A roadmap for understanding the evolutionary significance of structural genomic variation. Trends Ecol. Evol. 35(7): 561-572.

Meuwissen, T. H., B. J. Hayes and M. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157(4): 1819-1829.

Mikheenko, A., A. Prjibelski, V. Saveliev, D. Antipov and A. Gurevich. 2018. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics 34(13): i142-i150.

Nguyen, L.-T., H. A. Schmidt, A. von Haeseler and B. Q. Minh. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32(1): 268-274.

Nunn, A., I. Rodríguez-Arévalo, Z. Tandukar, K. Frels, A. Contreras-Garrido, P. Carbonell-Bejerano, P. Zhang, D. Ramos Cruz, K. Jandrasits and C. Lanz. 2022. Chromosome-level Thlaspi arvense genome provides new tools for translational research and for a newly domesticated cash cover crop of the cooler climates. Plant Biotechnol. J. 20(5): 944-963.

Oxley, P. R., M. Spivak and B. P. Oldroyd. 2010. Six quantitative trait loci influence task thresholds for hygienic behaviour in honeybees (*Apis mellifera*). Mol. Ecol. 19(7): 1452-1461.

Park, D., J. W. Jung, B.-S. Choi, M. Jayakodi, J. Lee, J. Lim, Y. Yu, Y.-S. Choi, M.-L. Lee and Y. Park. 2015. Uncovering the novel characteristics of Asian honey bee, *Apis cerana*, by whole genome sequencing. BMC Genomics 16(1): 1.

Potts, S. G., J. C. Biesmeijer, C. Kremen, P. Neumann, O. Schweiger and W. E. Kunin. 2010. Global pollinator declines: trends, impacts and drivers. Trends Ecol. Evol. 25(6): 345-353.

Radloff, S. E., C. Hepburn, H. Randall Hepburn, S. Fuchs, S. Hadisoesilo, K. Tan, M. S. Engel and V. Kuznetsov. 2010. Population structure and classification of *Apis cerana*. Apidologie 41(6): 589-601.

Ruttner, F. 2013. Biogeography and taxonomy of honeybees. Springer Science & Business Media.

Sayers, E. W., J. Beck, E. E. Bolton, J. R. Brister, J. Chan, R. Connor, M. Feldgarden, A. M. Fine, K. Funk and J. Hoffman. 2025. Database resources of the National Center for Biotechnology Information in 2025. Nucleic Acids Res. 53(D1): D20-D29.

Seppey, M., M. Manni and E. V. Zdobnov. 2019. BUSCO: Assessing genome assembly and annotation completeness. Methods Mol. Biol. 1962: 227-245.

Spötter, A., P. Gupta, G. Nürnberg, N. Reinsch and K. Bienefeld. 2012. Development of a 44K SNP assay focussing on the analysis of a varroa-specific defence behaviour in honey bees (*Apis mellifera carnica*). Mol. Ecol. Resour. 12(2): 323-332.

Takahashi, J.-i., T. Wakamiya, T. Kiyoshi, H. Uchiyama, S. Yajima, K. Kimura and T. Nomura. 2016. The complete mitochondrial genome of the Japanese honeybee, *Apis cerana japonica* (Insecta: Hymenoptera: Apidae). Mitochondrial DNA Part B 1(1): 156-157.

Tan, K., Z. Wang, H. Li, S. Yang, Z. Hu, G. Kastberger and B. P. Oldroyd. 2012. An 'I see you'prey-predator signal between the Asian honeybee, *Apis cerana*, and the hornet, *Vespa velutina*. Anim. Behav. 83(4): 879-882.

Uzunov, A., E. Brascamp and R. Büchler. 2017. The basic concept of honey bee breeding programs. Bee World 94(3): 84-87.

Wallberg, A., I. Bunikis, O. V. Pettersson, M.-B. Mosbech, A. K. Childers, J. D. Evans, A. S. Mikheyev, H. M. Robertson, G. E. Robinson and M. T. Webster. 2019. A hybrid *de novo* genome assembly of the honeybee, *Apis mellifera*, with chromosome-length scaffolds. BMC Genomics 20(1): 275.

Wallberg, A., F. Han, G. Wellhagen, B. Dahle, M. Kawata, N. Haddad, Z. L. P. Simões, M. H. Allsopp, I. Kandemir and P. De la Rúa. 2014. A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. Nat. Genet. 46(10): 1081-1088.

Wallberg, A., C. Schoening, M. T. Webster and M. Hasselmann. 2017. Two extended haplotype blocks are associated with adaptation to high altitude habitats in East African honey bees. PLoS Genet. 13(5): e1006792.

Wellenreuther, M. and L. Bernatchez. 2018. Eco-evolutionary genomics of chromosomal inversions. Trends Ecol. Evol. 33(6): 427-440.

Wenger, A. M., P. Peluso, W. J. Rowell, P.-C. Chang, R. J. Hall, G. T. Concepcion, J. Ebler, A. Fungtammasan, A. Kolesnikov and N. D. Olson. 2019. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat. Biotechnol. 37(10): 1155-1162.

Wragg, D., M. Marti-Marimon, B. Basso, J.-P. Bidanel, E. Labarthe, O. Bouchez, Y. Le Conte and A. Vignal. 2016. Whole-genome resequencing of honeybee drones to detect genomic selection in a population managed for royal jelly. Sci. Rep. 6(1): 27168.

SUPPLEMENTARY MATERIALS

Supplementary Table 1. Summary of PacBio HiFi sequencing raw data metrics for *Apis mellifera* and *Apis cerana* strains

NABIC ID	Strain	HiFi reads	HiFi yield (bp)	Read length N50 (bp)	Read quality (median)	Quality score base percentage		
						≥Q20	≥Q30	≥Q40
Apis mellifera								
DL000153	Am-A	1,064,338	12,211,262,054	13,981	Q38	100%	78%	43%
DL000160	Am-C	777,008	10,482,485,727	17,038	Q36	100%	75%	37%
DL000161	Am-D	944,251	10,350,837,510	13,169	Q39	100%	79%	45%
DL000162	Am-F	1,066,171	12,450,925,843	14,184	Q38	100%	78%	43%
DL000163	Am-V	1,116,667	11,780,254,254	12,633	Q39	100%	80%	46%
Apis cerana								
DL000166	Ac-R	1,115,798	11,549,704,693	10,247	Q39	100%	80%	46%
DL000199	Ac-X	959,376	11,807,852,449	15,638	Q37	100%	69%	28%

Supplementary Table 2. Chromosome-level scaffold lengths generated *via* Pore-C (Hi-C) proximity ligation mapping

Scaffold ID	Am-A (bp)	Am-C (bp)	Am-D (bp)	Am-F (bp)	Am-V (bp)	Ac-R (bp)	Ac-X (bp)
Scaffold_1	27,781,000	27,791,684	27,785,862	27,782,125	27,754,436	21,052,746	26,973,790
Scaffold_2	17,325,499	17,823,053	17,423,578	17,807,206	17,798,316	20,591,105	17,173,956
Scaffold_3	16,369,877	16,480,392	16,417,924	16,456,325	16,403,597	18,007,874	15,848,518
Scaffold_4	16,276,586	16,350,082	16,271,449	16,231,972	16,180,488	15,398,461	15,748,066
Scaffold_5	14,293,130	14,265,236	14,260,943	14,027,659	14,278,403	13,459,284	13,610,906
Scaffold_6	13,998,226	14,016,114	13,910,630	13,983,796	14,012,453	12,912,559	13,472,413
Scaffold_7	13,620,572	13,605,269	13,584,843	13,588,285	13,576,834	12,789,988	13,181,083
Scaffold_8	13,482,798	13,497,867	13,473,980	13,488,223	13,539,242	11,883,068	13,130,291
Scaffold_9	12,708,511	12,743,519	12,744,634	12,747,928	12,711,660	11,636,894	12,144,468
Scaffold_10	12,551,974	12,597,727	12,664,734	12,524,161	12,569,578	11,455,735	12,099,923
Scaffold_11	12,346,458	12,414,409	12,347,062	12,340,733	12,338,885	11,005,288	11,896,615
Scaffold_12	11,613,761	11,556,718	11,552,112	11,535,876	11,532,450	10,623,735	11,851,483
Scaffold_13	11,511,350	11,477,876	11,422,416	11,396,078	11,434,075	10,390,644	11,138,990
Scaffold_14	11,417,804	11,440,874	11,310,771	11,113,873	11,415,776	10,269,928	10,625,668
Scaffold_15	9,532,013	9,556,725	9,591,882	9,516,440	9,508,215	9,148,161	9,454,152
Scaffold_16	7,587,657	7,606,986	7,634,152	7,603,339	7,612,954	7,058,583	7,260,788
Total length	222,417,216	223,224,531	222,396,972	222,144,019	222,667,362	207,684,053	215,611,110

Supplementary Table 3. Species-level comparative enrichment analysis of Gene Ontology (GO) Level 2 categories between *Apis cerana* (n = 2) and *Apis mellifera* (n = 5) based on Fisher's exact test on pooled absolute gene counts.

Category	Term	Ac_ count	Am_ count	Ac_pct	Am_pct	Difference_ pct	p_value_ fisher	p_value_ chi2	odds_ ratio	adj_p_ fisher	adj_p_ chi2	Significant_ Fisher	Significant_ Chi2
Molecular function	Catalytic activity	4911	16253	39.56974	44.991	−5.42127	6.5E-26	8.94E-26	0.800602	2.79E-24	3.85E-24	TRUE	TRUE
Molecular function	Transporter activity	876	3309	7.058255	9.159862	−2.10161	2.33E-13	7.1E-13	0.753139	3.34E-12	1.02E-11	TRUE	TRUE
Cellular component	Cellular anatomical structure	10494	29506	84.55402	81.67751	2.876516	2.33E-13	4.22E-13	1.228007	3.34E-12	9.08E-12	TRUE	TRUE
Molecular function	ATP-dependent activity	614	2214	4.947224	6.12872	−1.1815	9.04E-07	1.4E-06	0.797186	9.71E-06	1.5E-05	TRUE	TRUE
Cellular component	Protein-containing complex	2908	7736	23.43083	21.41453	2.016295	3.28E-06	3E-06	1.122968	2.82E-05	2.58E-05	TRUE	TRUE
Biological process	Multicellular organismal process	740	1785	5.962453	4.941176	1.021276	1.3E-05	1.1E-05	1.219792	9.33E-05	7.88E-05	TRUE	TRUE
Biological process	Biological regulation	3672	10122	29.58666	28.01938	1.56728	0.000894	0.000874	1.079439	0.005489	0.005369	TRUE	TRUE
Biological process	Developmental process	905	2331	7.291918	6.452595	0.839323	0.001318	0.001314	1.140306	0.007086	0.007061	TRUE	TRUE
Biological process	Regulation of biological process	3503	9672	28.22496	26.7737	1.451259	0.001787	0.001779	1.07552	0.008536	0.008501	TRUE	TRUE
Biological process	Localization	2150	6655	17.32334	18.42215	−1.0988	0.006133	0.006391	0.927857	0.024102	0.024984	TRUE	TRUE
Biological process	Positive regulation of biological process	706	1824	5.688502	5.049135	0.639367	0.006166	0.006123	1.134267	0.024102	0.024984	TRUE	TRUE
Molecular function	Molecular adaptor activity	341	841	2.747563	2.328028	0.419535	0.010275	0.009819	1.185302	0.03682	0.035184	TRUE	TRUE
Molecular function	Antioxidant activity	48	199	0.386754	0.550865	−0.16411	0.028023	0.032067	0.700928	0.09269	0.100978	FALSE	FALSE
Biological process	Negative regulation of biological process	688	1823	5.54347	5.046367	0.497103	0.032545	0.032877	1.104288	0.099958	0.100978	FALSE	FALSE
Molecular function	Molecular transducer activity	376	963	3.029571	2.665744	0.363827	0.035993	0.03545	1.140746	0.103181	0.101623	FALSE	FALSE
Biological process	Immune system process	109	252	0.878253	0.697578	0.180675	0.045701	0.049939	1.261299	0.122821	0.134211	FALSE	FALSE
Molecular function	Electron transfer activity	37	153	0.298123	0.423529	−0.12541	0.055288	0.064763	0.703015	0.139846	0.163811	FALSE	FALSE
Biological process	Response to stimulus	1905	5375	15.34929	14.87889	0.470394	0.20499	0.210716	1.037347	0.489697	0.503376	FALSE	FALSE
Molecular function	Cytoskeletal motor activity	103	263	0.829909	0.728028	0.101881	0.253725	0.28383	1.141113	0.562194	0.599146	FALSE	FALSE
Biological process	Reproductive process	185	489	1.490613	1.353633	0.13698	0.266374	0.279879	1.102726	0.562194	0.599146	FALSE	FALSE
Biological process	Detoxification	48	168	0.386754	0.465052	−0.0783	0.27456	0.292606	0.830982	0.562194	0.599146	FALSE	FALSE
Molecular function	Molecular function regulator activity	687	1913	5.535412	5.295502	0.23991	0.309346	0.316849	1.047959	0.604631	0.619295	FALSE	FALSE
Molecular function	Protein-containing complex destabilizing activity	6	12	0.048344	0.033218	0.015126	0.426389	0.627769	1.455582	0.797162	0.975935	FALSE	FALSE

Supplementary Table 3. Continued

Category	Term	Ac_ count	Am_ count	Ac_pct	Am_pct	Difference_ pct	p_value_ fisher	p_value_ chi2	odds_ ratio	adj_p_ fisher	adj_p_ chi2	Significant_ Fisher	Significant_ Chi2
Molecular function	Molecular carrier activity	28	95	0.225606	0.262976	−0.03737	0.535111	0.541268	0.857576	0.950022	0.975935	FALSE	FALSE
Molecular function	Toxin activity	0	4	0	0.011073	−0.01107	0.578253	0.54901	0	0.950022	0.975935	FALSE	FALSE
Biological process	Rhythmic process	21	53	0.169205	0.146713	0.022492	0.594083	0.673962	1.153566	0.950022	0.975935	FALSE	FALSE
Molecular Function	Translation regulator activity	18	47	0.145033	0.130104	0.014929	0.671274	0.802517	1.114912	0.950022	0.975935	FALSE	FALSE
Molecular Function	Translation factor activity	165	463	1.329466	1.281661	0.047805	0.678802	0.718451	1.037802	0.950022	0.975935	FALSE	FALSE
Biological process	Locomotion	33	105	0.265893	0.290657	−0.02476	0.696947	0.726863	0.914572	0.950022	0.975935	FALSE	FALSE
Biological process	Pigmentation	10	25	0.080574	0.069204	0.01137	0.69902	0.831112	1.164422	0.950022	0.975935	FALSE	FALSE
Molecular function	Binding	7794	22755	62.79913	62.98962	−0.19049	0.706177	0.712642	0.991871	0.950022	0.975935	FALSE	FALSE
Biological process	Growth	27	73	0.217549	0.202076	0.015473	0.731387	0.831145	1.076736	0.950022	0.975935	FALSE	FALSE
Biological process	Cellular process	9425	27481	75.9407	76.07197	−0.13127	0.769906	0.776873	0.992827	0.950022	0.975935	FALSE	FALSE
Molecular function	Cargo receptor activity	17	46	0.136975	0.127336	0.00964	0.773756	0.910163	1.075806	0.950022	0.981498	FALSE	FALSE
Molecular function	Protein folding chaperone	40	110	0.322295	0.304498	0.017796	0.778591	0.830212	1.058634	0.950022	0.975935	FALSE	FALSE
Biological process	Biological process involved in interspecies interaction between organisms	81	245	0.652647	0.678201	−0.02555	0.799114	0.812658	0.962074	0.950022	0.975935	FALSE	FALSE
Molecular function	General transcription initiation factor activity	25	79	0.201434	0.218685	−0.01725	0.822026	0.805636	0.920956	0.950022	0.975935	FALSE	FALSE
Molecular function	Transcription regulator activity	898	2634	7.235517	7.291349	−0.05583	0.85696	0.85201	0.991745	0.950022	0.975935	FALSE	FALSE
Biological process	Viral process	10	33	0.080574	0.091349	−0.01078	0.861648	0.862454	0.881943	0.950022	0.975935	FALSE	FALSE
Molecular function	Structural molecule activity	330	952	2.658932	2.635294	0.023637	0.896753	0.913021	1.009215	0.96401	0.981498	FALSE	FALSE
Biological process	Homeostatic process	186	545	1.498671	1.508651	−0.00998	0.965942	0.97125	0.993284	1	1	FALSE	FALSE
Molecular function	Nutrient reservoir activity	6	18	0.048344	0.049827	−0.00148	1	1	0.970227	1	1	FALSE	FALSE
Cellular component	Virion component	2	5	0.016115	0.013841	0.002274	1	1	1.164316	1	1	FALSE	FALSE

For each of the 44 evaluated GO Level 2 terms across the three primary ontologies (Biological Process, Molecular Function, and Cellular Component), the table displays the pooled absolute gene counts, relative percentages (%) in the respective gene space of each species, relative difference in percentage (*A. cerana* % - *A. mellifera* %), raw *p*-values from Fisher's exact test, and adjusted *p*-values under Benjamini-Hochberg False Discovery Rate (FDR) control. Statistically significant enrichment is defined at an adjusted *p*-value < 0.05.

Supplementary Table 4. Summary of assembled mitochondrial genome characteristics and nucleotide compositions for seven Korean honeybee strains

Species/Strain	Length (bp)	A (%)	T (%)	G (%)	C (%)	AT content (%)	GC content (%)
Apis cerana							
Ac-R	16,224	42.47	41.82	6.14	9.57	84.29	15.71
Ac-X	16,247	42.47	41.86	6.15	9.52	84.33	15.67
Apis mellifera							
Am-A	16,576	43.34	41.75	5.45	9.46	85.09	14.91
Am-C	16,515	43.28	41.79	5.47	9.46	85.07	14.93
Am-D	16,539	43.32	41.75	5.48	9.45	85.07	14.93
Am-F	16,555	43.32	41.76	5.45	9.47	85.07	14.93
Am-V	16,546	43.33	41.75	5.48	9.45	85.08	14.92

Supplementary Table 5. Complete list of the 65 mitochondrial genomes used in the phylogenomic reanalysis, including taxonomic classifications, GenBank accession numbers, countries of origin, and source categories

No.	GenBank accession	Taxonomic species	Strain/Breeding line	Country of origin	Source category
1	This study (Ac-R)	*Apis cerana*	Ac-R	South Korea	This study (National seed bank)
2	This study (Ac-X)	*Apis cerana*	Ac-X	South Korea	This study (National seed bank)
3	This study (Am-A)	*Apis mellifera*	Am-A	South Korea	This study (National seed bank)
4	This study (Am-C)	*Apis mellifera*	Am-C	South Korea	This study (National seed bank)
5	This study (Am-D)	*Apis mellifera*	Am-D	South Korea	This study (National seed bank)
6	This study (Am-F)	*Apis mellifera*	Am-F	South Korea	This study (National seed bank)
7	This study (Am-V)	*Apis mellifera*	Am-V	South Korea	This study (National seed bank)
8	NC_039709	*Apis andreniformis*	–	Thailand	NCBI reference database
9	AP017942	*Apis cerana*	–	China	NCBI reference database
10	AP017983	*Apis cerana*	–	China	NCBI reference database
11	AP018368	*Apis cerana*	–	China	NCBI reference database
12	OP689704	*Apis cerana*	–	China	NCBI reference database
13	OR897030	*Apis cerana*	–	China	NCBI reference database
14	OR897032	*Apis cerana*	–	China	NCBI reference database
15	OR897033	*Apis cerana*	–	China	NCBI reference database
16	PP692293	*Apis cerana*	–	China	NCBI reference database
17	PV088921	*Apis cerana*	–	India	NCBI reference database
18	AP017985	*Apis cerana*	–	Japan	NCBI reference database
19	GQ162109	*Apis cerana*	–	South Korea	NCBI reference database
20	KX908206	*Apis cerana*	–	South Korea	NCBI reference database
21	OR936096	*Apis cerana*	–	South Korea	NCBI reference database
22	PP175370	*Apis cerana*	–	South Korea	NCBI reference database
23	PP175371	*Apis cerana*	–	South Korea	NCBI reference database
24	NC_037709	*Apis dorsata*	–	India	NCBI reference database
25	AP018369	*Apis dorsata*	–	Thailand	NCBI reference database
26	AP018491	*Apis florea*	–	Thailand	NCBI reference database
27	NC_021401	*Apis florea*	–	Thailand	NCBI reference database
28	AP017643	*Apis koschevnikovi*	–	Malaysia	NCBI reference database
29	OP764682	*Apis laboriosa*	–	China	NCBI reference database
30	OP764683	*Apis laboriosa*	–	China	NCBI reference database

Supplementary Table 5. Continued

No.	GenBank accession	Taxonomic species	Strain/Breeding line	Country of origin	Source category
31	NC_036155	*Apis laboriosa*	–	Nepal	NCBI reference database
32	KM458618	*Apis mellifera*	–	Algeria	NCBI reference database
33	OM203296	*Apis mellifera*	–	China	NCBI reference database
34	OM203319	*Apis mellifera*	–	China	NCBI reference database
35	OM203321	*Apis mellifera*	–	China	NCBI reference database
36	OM203329	*Apis mellifera*	–	China	NCBI reference database
37	OM203348	*Apis mellifera*	–	China	NCBI reference database
38	KY464958	*Apis mellifera*	–	Egypt	NCBI reference database
39	MN585108	*Apis mellifera*	–	Ethiopia	NCBI reference database
40	KY614238	*Apis mellifera*	–	Germany	NCBI reference database
41	MZ981768	*Apis mellifera*	–	Italy	NCBI reference database
42	AP018432	*Apis mellifera*	–	Japan	NCBI reference database
43	AP018434	*Apis mellifera*	–	Japan	NCBI reference database
44	AP018435	*Apis mellifera*	–	Japan	NCBI reference database
45	MF678581	*Apis mellifera*	–	Kenya	NCBI reference database
46	MN119925	*Apis mellifera*	–	Madagascar	NCBI reference database
47	MN714162	*Apis mellifera*	–	Malta	NCBI reference database
48	MN585109	*Apis mellifera*	–	Nigeria	NCBI reference database
49	MN250878	*Apis mellifera*	–	Slovenia	NCBI reference database
50	KX870183	*Apis mellifera*	–	South Africa	NCBI reference database
51	MG552689	*Apis mellifera*	–	South Africa	NCBI reference database
52	MG552691	*Apis mellifera*	–	South Africa	NCBI reference database
53	MG552694	*Apis mellifera*	–	South Africa	NCBI reference database
54	MG552698	*Apis mellifera*	–	South Africa	NCBI reference database
55	MG552703	*Apis mellifera*	–	South Africa	NCBI reference database
56	OK075087	*Apis mellifera*	–	Spain	NCBI reference database
57	NC_001566	*Apis mellifera*	–	Sweden	NCBI reference database
58	KP163643	*Apis mellifera*	–	Syria	NCBI reference database
59	MT188686	*Apis mellifera*	–	Turkey	NCBI reference database
60	MN714161	*Apis mellifera*	–	Yemen	NCBI reference database
61	AP018398	*Apis nigrocincta*	–	Indonesia	NCBI reference database
62	NC_038114	*Apis nigrocincta*	–	Indonesia	NCBI reference database
63	AP018157	*Apis nuluensis*	–	Malaysia	NCBI reference database
64	NC_036235	*Apis nuluensis*	–	Malaysia	NCBI reference database
65	DQ870926	*Bombus ignitus*	–	Japan	NCBI reference database

Supplementary Table 6. Detailed annotation profile of predicted gene models assigned to the Gene Ontology (GO) term 'toxin activity' (GO:0090303) across the seven honeybee genomes

Strain	SeqID	Description	Blast_Top_Hit	GO_IDs	GO_Names
Am-A	g1420.t1	XP_006566787. 126S proteasome non-ATPase regulatory subunit 8	XP_016908963.1 26S proteasome non-ATPase regulatory subunit 8 [Apis cerana]/ XP_006566787.1 26S proteasome non-ATPase regulatory subunit 8 [Apis mellifera]/ PBC32574.1 26S proteasome non-ATPase regulatory subunit [Apis cerana cerana]	GO:0005634| GO:0005829| GO:0008541| GO:0035821| GO:0043161| GO:0090729	Nucleus|cytosol|proteasome r egulatory particle, lid subcomplex|modulation of process of another organism|proteasome-mediated ubiquitin-dependent protein catabolic process|toxin activity
Am-A	g2757.t1	XP_006560077.1 omega-conotoxin-like protein 1	XP_006560077.1 omega-conotoxin-like protein 1 [Apis mellifera]/ AHZ46208.1 inhibitor cysteine knot peptide [Apis cerana]/ XP_016917762.1 omega-conotoxin-like protein 1 [Apis cerana]/ H9KQJ7.1 RecName: Full = Omega-conotoxin-like protein 1; Short = OCLP1; Flags: Precursor	GO:0035821| GO:0042742| GO:0090729| GO:0099106	Modulation of process of another organism|defense response to bacterium|toxin activity|ion channel regulator activity
Am-A	g6923.t1	WP_046325090.1 MFS transporter	WP_046325090.1 MFS transporter [Lactobacillus melliventris]/ KJY56835.1 Transporter-membrane protein [Lactobacillus melliventris]	GO:0005886| GO:0008200| GO:0022857| GO:0042151| GO:0055085| GO:0090729	plasma membrane|ion channel inhibitor activity|transmembrane transporter activity|nematocyst|transmembrane transport|toxin activity
Am-C	g1408.t1	XP_006566787.1 26S proteasome non-ATPase regulatory subunit 8	XP_016908963.1 26S proteasome non-ATPase regulatory subunit 8 [Apis cerana]/ XP_006566787.1 26S proteasome non-ATPase regulatory subunit 8 [Apis mellifera]/ PBC32574.1 26S proteasome non-ATPase regulatory subunit [Apis cerana cerana]	GO:0005634| GO:0005829| GO:0008541| GO:0035821| GO:0043161| GO:0090729	nucleus|cytosol|proteasome regulatory particle, lid subcomplex|modulation of process of another organism|proteasome-mediated ubiquitin-dependent protein catabolic process|toxin activity
Am-C	g3091.t1	PBC31321.1 hypothetical protein APICC_05944	PBC31321.1 hypothetical protein APICC_05944 [Apis cerana cerana]	GO:0005615| GO:0008970| GO:0016020| GO:0016042| GO:0035821| GO:0042742| GO:0090729| GO:0099106	extracellular space|phospholipase A1 activity|membrane|lipid catabolic process|modulation of process of another organism|defense response to bacterium|toxin activity|ion channel regulator activity
Am-D	g1795.t1	XP_006566787.1 26S proteasome non-ATPase regulatory subunit 8	XP_016908963.1 26S proteasome non-ATPase regulatory subunit 8 [Apis cerana]/ XP_006566787.1 26S proteasome non-ATPase regulatory subunit 8 [Apis mellifera]/ PBC32574.1 26S proteasome non-ATPase regulatory subunit [Apis cerana cerana]	GO:0005634| GO:0005829| GO:0008541| GO:0035821| GO:0043161| GO:0090729	nucleus|cytosol|proteasome regulatory particle, lid subcomplex|modulation of process of another organism|proteasome-mediated ubiquitin-dependent protein catabolic process|toxin activity

Supplementary Table 6. Continued

Strain	SeqID	Description	Blast_Top_Hit	GO_IDs	GO_Names
Am-F	g1420.t1	XP_006566787.1 26S proteasome non-ATPase regulatory subunit 8	XP_016908963.1 26S proteasome non-ATPase regulatory subunit 8 [Apis cerana]/ XP_006566787.1 26S proteasome non-ATPase regulatory subunit 8 [Apis mellifera]/ PBC32574.1 26S proteasome non-ATPase regulatory subunit [Apis cerana cerana]	GO:0005634| GO:0005829| GO:0008541| GO:0035821| GO:0043161| GO:0090729	nucleus|cytosol|proteasome regulatory particle, lid subcomplex|modulation of process of another organism|proteasome-mediated ubiquitin-dependent protein catabolic process|toxin activity
Am-F	g2742.t1	XP_006560077.1 omega-conotoxin-like protein 1	XP_006560077.1 omega-conotoxin-like protein 1 [Apis mellifera]/ AHZ46208.1 inhibitor cysteine knot peptide [Apis cerana]/XP_016917762.1 omega-conotoxin-like protein 1 [Apis cerana]/H9KQJ7.1 RecName: Full = Omega-conotoxin-like protein 1; Short = OCLP1; Flags: Precursor	GO:0035821| GO:0042742| GO:0090729| GO:0099106	modulation of process of another organism|defense response to bacterium|toxin activity|ion channel regulator activity
Am-V	g1797.t1	XP_006566787.1 26S proteasome non-ATPase regulatory subunit 8	XP_016908963.1 26S proteasome non-ATPase regulatory subunit 8 [Apis cerana]/ XP_006566787.1 26S proteasome non-ATPase regulatory subunit 8 [Apis mellifera]/ PBC32574.1 26S proteasome non-ATPase regulatory subunit [Apis cerana cerana]	GO:0005634| GO:0005829| GO:0008541| GO:0035821| GO:0043161| GO:0090729	nucleus|cytosol|proteasome regulatory particle, lid subcomplex|modulation of process of another organism|proteasome-mediated ubiquitin-dependent protein catabolic process|toxin activity
Am-V	g2757.t1	XP_006560077.1 omega-conotoxin-like protein 1	XP_006560077.1 omega-conotoxin-like protein 1 [Apis mellifera]/ AHZ46208.1 inhibitor cysteine knot peptide [Apis cerana]/ XP_016917762.1 omega-conotoxin-like protein 1 [Apis cerana]/H9KQJ7.1 RecName: Full = Omega-conotoxin-like protein 1; Short = OCLP1; Flags: Precursor	GO:0035821| GO:0042742| GO:0090729| GO:0099106	modulation of process of another organism|defense response to bacterium|toxin activity|ion channel regulator activity
Am-V	g8121.t1	WP_084547134.1 hypothetical protein	ORF16435.1 hypothetical protein BGI01_00410 [Snodgrassella alvi]/ ORF09532.1 hypothetical protein BGH98_00715 [Snodgrassella alvi]/ ORF22040.1 hypothetical protein BGI03_00695 [Snodgrassella alvi]/ WP_084547134.1hypothetical protein [Snodgrassella alvi]/ORF22446.1 hypothetical protein BGI04_00900 [Snodgrassella alvi]	GO:0005509| GO:0005576| GO:0016020| GO:0035821| GO:0090729	calcium ion binding|extracellular region|membrane|modulation of process of another organism|toxin activity

Supplementary Table 6. Continued

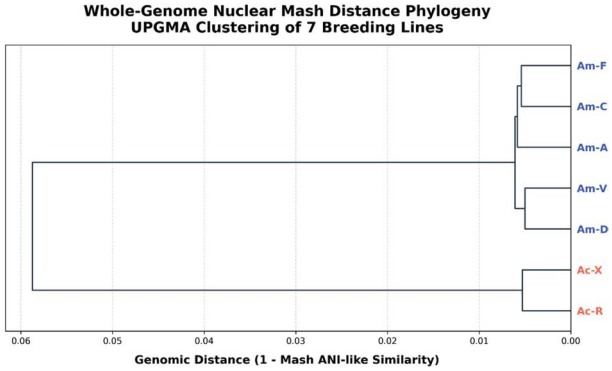
Strain	SeqID	Description	Blast_Top_Hit	GO_IDs	GO_Names
Am-V	g8153.t1	WP_084554909.1 hypothetical protein	WP_084554909.1hypothetical protein [Snodgrassella alvi]/ORF24010.1 hypothetical protein BGI07_08820 [Snodgrassella alvi]/ORF33967.1 hypothetical protein BGI10_00830 [Snodgrassella alvi]/ORF34327.1 hypothetical protein BGI11_05740 [Snodgrassella alvi]/ORF43338.1 hypothetical protein BGI15_04400 [Snodgrassella alvi]/ORF40944.1 hypothetical protein BGI14_04050 [Snodgrassella alvi]	GO:0004035| GO:0005509| GO:0005576| GO:0016020| GO:0035821| GO:0090729	alkaline phosphatase activity|calcium ion binding|extracellular region|membrane|modulation of process of another organism|toxin activity

The table lists the isolate strain, gene identifier (SeqID), consensus protein description, top BLAST hit against the NCBI non-redundant (nr) protein database, mapped GO terms/accessions, and the biological classification of each candidate gene. The classification distinguishes host-specific eukaryotic defense peptides (e.g., omega-conotoxin-like protein 1/inhibitor cysteine knot peptide) from bacterial sequence contaminants derived from co-sequenced gut symbionts (*Snodgrassella alvi* or *Lactobacillus*).

Supplementary Table 7. Comparison of Cytochrome P450 (CYP) monooxygenase gene copy numbers across the seven chromosome-level honeybee genome assemblies

Strain	Species	P450_Count
Ac-R	*A. cerana*	18
Ac-X	*A. cerana*	18
Am-A	*A. mellifera*	19
Am-C	*A. mellifera*	21
Am-D	*A. mellifera*	20
Am-F	*A. mellifera*	21
Am-V	*A. mellifera*	21

The table summarizes the total counts of identified Cytochrome P450 genes for the two *Apis cerana* isolates (Ac-R, Ac-X) and the five *Apis mellifera* isolates (Am-A, Am-C, Am-D, Am-F, Am-V) based on sequence homology matching against the NCBI nr database and functional domain signatures (InterProScan).



Supplementary Fig. 1. Whole-genome nuclear Mash distance phylogeny tree of the seven honeybee breeding lines. The UPGMA tree was constructed using Mash distances (k = 21, sketch size = 10,000) based on the chromosome-scale assembly fasta files. Scale bar indicates the genomic distance (1-Mash similarity). *A. cerana* (coral) and *A. mellifera* (blue) strains form distinct monophyletic groups with minimal intra-species genetic distance, confirming high nuclear genome integrity.